AGU ADVANCING EARTH AND SPACE SCIENCES

Earth and Space Science

RESEARCH ARTICLE
10.1029/2024EA003792

Aura Ozone Monitoring Instrument (OMI) Collection 4 Formaldehyde Products

Zolal Ayazpour[1,2], Gonzalo González Abad[1], Caroline R. Nowlan[1], Kang Sun[2,3], Hyeong-Ahn Kwon[1,4], Christopher Chan Miller[1,5,6,7], Heesung Chong[1], Huiqun Wang[1], Xiong Liu[1], Kelly Chance[1], Ewan O'Sullivan[1], Lei Zhu[8], Corinne Vigouroux[9], Isabelle De Smedt[9], Wolfgang Stremme[10], James W. Hannigan[11], Justus Notholt[12], Xiaoyu Sun[12], Mathias Palm[12], Cristof Petri[12], Kimberly Strong[13], Amelie N. Röhling[14], Emmanuel Mahieu[15], Dan Smale[16], Yao Té[17], Isamu Morino[18], Isao Murata[19], Tomoo Nagahama[20], Rigel Kivi[21], Maria Makarova[22], Nicholas Jones[23], Ralf Sussmann[24], and Minqiang Zhou[25]

[1]Center for Astrophysics|Harvard & Smithsonian, Cambridge, MA, USA, [2]Department of Civil, Structural and Environmental Engineering, University at Buffalo, Buffalo, NY, USA, [3]Research and Education in Energy, Environment and Water Institute, University at Buffalo, Buffalo, NY, USA, [4]Now at University of Suwon, Hwaseong, South Korea, [5]Harvard John A. Paulson School of Engineering and Applied Sciences, Harvard University, Cambridge, MA, USA, [6]Environmental Defense Fund, New York, NY, USA, [7]Climate Change Research Centre, University of New South Wales, Kensington, NSW, Australia, [8]School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China, [9]Royal Belgian Institute for Space Aeronomy (BIRA-IASB), Brussels, Belgium, [10]Instituto de Ciencias de la Atmósfera y Cambio Climático, Universidad Nacional Autónoma de México, Mexico City, Mexico, [11]Atmospheric Chemistry, Observations & Modeling, National Center for Atmospheric Research (NCAR), Boulder, CO, USA, [12]Institute of Environmental Physics, University of Bremen, Bremen, Germany, [13]Department of Physics, University of Toronto, Toronto, ON, Canada, [14]Institute of Meteorology and Climate Research (IMK-ASF), Karlsruhe Institute of Technology (KIT), Karlsruhe, Germany, [15]Institut d'Astrophysique et de Géophysique, Université de Liège, Liège, Belgium, [16]National Institute of Water & Atmospheric Research (NIWA), Lauder, New Zealand, [17]Sorbonne Université, CNRS, MONARIS, UMR8233, Paris, France, [18]National Institute for Environmental Studies (NIES), Tsukuba, Japan, [19]Graduate School of Environmental Studies, Tohoku University, Sendai, Japan, [20]Institute for Space-Earth Environmental Research (ISEE), Nagoya University, Nagoya, Japan, [21]Space and Earth Observation Centre, Finnish Meteorological Institute, Sodankylä, Finland, [22]Atmospheric Physics Department, Saint Petersburg State University, St. Petersburg, Russia, [23]Centre for Atmospheric Chemistry, University of Wollongong, Wollongong, NSW, Australia, [24]Karlsruhe Institute of Technology (KIT), IMK-IFU, Garmisch-Partenkirchen, Germany, [25]Institute of Atmospheric Physics, Chinese Academy of Sciences, Beijing, China

Abstract This study presents the ozone monitoring instrument (OMI) Collection 4 formaldehyde (HCHO) retrieval developed with the Smithsonian Astrophysical Observatory's (SAO) Making Earth System Data Records for Use in Research Environments (MEaSUREs) algorithm. The retrieval algorithm updates and makes improvements to the NASA operational OMI HCHO (OMI Collection 3 HCHO) algorithm, and has been transitioned to use OMI Collection 4 Level-1B radiances. This paper describes the updated retrieval algorithm and compares Collection 3 and Collection 4 data products. The OMI Collection 4 HCHO exhibits remarkably improved stability over time in comparison to the OMI Collection 3 HCHO product, with better precision and the elimination of artificial trends present in the Collection 3 during the later years of the mission. We validate the OMI Collection 4 HCHO data product using Fourier-Transform Infrared (FTIR) ground-based HCHO measurements. The climatological monthly averaged OMI Collection 4 HCHO vertical column densities (VCDs) agree well with the FTIR VCDs, with a correlation coefficient of 0.83, root-mean-square error (RMSE) of 2.98 × 10^{15} molecules cm^{-2}, regression slope of 0.79, and intercept of 8.21 × 10^{14} molecules cm^{-2}. Additionally, we compare the monthly averaged OMI Collection 4 HCHO VCDs to OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI HCHO VCDs in overlapping years for 12 geographic regions. This comparison demonstrates high correlation coefficients of 0.98 (OMPS Suomi NPP), 0.97 (OMPS NOAA-20), and 0.90 (TROPOMI).

Key Points:

- We present the ozone monitoring instrument (OMI) Collection 4 formaldehyde (HCHO) product, retrieved using the SAO retrieval algorithm
- OMI Collection 4 HCHO product is well-correlated with the ground-based Fourier-Transform Infrared Spectrometers (FTIR) HCHO measurements
- OMI Collection 4 HCHO provides a stable long-term data record and shows strong agreement with OMPS and TROPOMI HCHO products

Correspondence to:
G. González Abad,
ggonzalezabad@cfa.harvard.edu

Citation:
Ayazpour, Z., González Abad, G., Nowlan, C. R., Sun, K., Kwon, H.-A., Chan Miller, C., et al. (2025). Aura ozone monitoring instrument (OMI) Collection 4 formaldehyde products. *Earth and Space Science*, *12*, e2024EA003792. https://doi.org/10.1029/2024EA003792

Received 31 MAY 2024
Accepted 1 NOV 2024

Author Contributions:
Conceptualization: Zolal Ayazpour, Gonzalo González Abad, Caroline R. Nowlan
Data curation: Zolal Ayazpour
Formal analysis: Zolal Ayazpour
Funding acquisition: Gonzalo González Abad, Caroline R. Nowlan, Kang Sun, Huiqun Wang, Xiong Liu, Kelly Chance
Investigation: Zolal Ayazpour, Heesung Chong
Methodology: Zolal Ayazpour, Gonzalo González Abad, Caroline R. Nowlan
Project administration: Zolal Ayazpour, Gonzalo González Abad
Resources: Lei Zhu, Corinne Vigouroux, Isabelle De Smedt, Wolfgang Stremme, James W. Hannigan, Justus Notholt, Xiaoyu Sun, Mathias Palm, Cristof Petri, Kimberly Strong, Amelie N. Röhling, Emmanuel Mahieu, Dan Smale, Yao Té,

1. Introduction

Global measurements of formaldehyde (HCHO) columns from space provide a rich data set to support studies on atmospheric chemistry and air quality. Atmospheric HCHO is mainly concentrated in the troposphere. Local and

Isamu Morino, Isao Murata, Tomoo Nagahama, Rigel Kivi, Maria Makarova, Nicholas Jones, Ralf Sussmann, Minqiang Zhou
Software: Zolal Ayazpour, Gonzalo González Abad, Caroline R. Nowlan, Hyeong-Ahn Kwon, Christopher Chan Miller, Heesung Chong, Huiqun Wang, Xiong Liu, Kelly Chance, Ewan O'Sullivan
Supervision: Gonzalo González Abad, Caroline R. Nowlan, Kang Sun, Xiong Liu, Kelly Chance
Validation: Zolal Ayazpour, Gonzalo González Abad, Caroline R. Nowlan, Hyeong-Ahn Kwon
Visualization: Zolal Ayazpour
Writing – original draft: Zolal Ayazpour
Writing – review & editing: Zolal Ayazpour, Gonzalo González Abad, Caroline R. Nowlan, Kang Sun, Hyeong-Ahn Kwon, Christopher Chan Miller, Heesung Chong, Huiqun Wang, Xiong Liu, Kelly Chance, Ewan O'Sullivan, Lei Zhu, Corinne Vigouroux, Isabelle De Smedt, Wolfgang Stremme, James W. Hannigan, Justus Notholt, Xiaoyu Sun, Mathias Palm, Cristof Petri, Kimberly Strong, Amelie N. Röhling, Emmanuel Mahieu, Dan Smale, Yao Té, Isamu Morino, Isao Murata, Tomoo Nagahama, Rigel Kivi, Maria Makarova, Nicholas Jones, Ralf Sussmann

regional enhanced concentrations of HCHO are primarily due to oxidation of non-methane volatile organic compounds (NMVOCs) from biogenic sources, anthropogenic activities, and biomass burning, while the global HCHO background concentration originates primarily from oxidation of methane. Given the influence of NMVOCs on hydroxyl radical (OH) concentration (Valin et al., 2016; Wolfe et al., 2019), tropospheric ozone (O_3) formation (Jin et al., 2017; Martin et al., 2004; Souri et al., 2020; Travis et al., 2022), and secondary organic aerosols production (Liao et al., 2019; Marais et al., 2016; Veefkind et al., 2011), NMVOCs play an important role in tropospheric composition (Anderson et al., 1996). Since most NMVOC emissions cannot be directly observed from space, formaldehyde measurements from satellites serve as a proxy for overall NMVOC emissions to estimate top-down isoprene emission (Barkley et al., 2013; Kaiser et al., 2018; Marais et al., 2012; Millet et al., 2008; Stavrakou et al., 2009a, 2009b) thanks to formaldehyde's short lifetime (Brune et al., 1999).

Ultraviolet (UV) nadir instruments can detect tropospheric HCHO from space. In the past few decades, several Low Earth Orbit (LEO) satellites have been launched to make hyperspectral measurements for deriving atmospheric composition. Thomas et al. (1998) and Chance et al. (2000) used Global Ozone Monitoring Experiment (GOME) satellite measurements to produce the first global HCHO product. Since then, different retrieval algorithms have been used for various satellite instruments to extract HCHO column concentrations, including from the Scanning Imaging Absorption spectrometer for Atmospheric CHartography (SCIAMACHY) (De Smedt et al., 2008; Wittrock et al., 2006), GOME-2 (De Smedt et al., 2012, 2015; Vrekoussis et al., 2010), ozone monitoring instrument (OMI) (De Smedt et al., 2015; González Abad et al., 2015; Kurosu et al., 2004), Ozone Mapping and Profiler Suite (OMPS) on the Suomi NPP (National Polar-orbiting Partnership) satellite (González Abad et al., 2016; Kwon et al., 2023; C. Li et al., 2015; Nowlan et al., 2023; Su et al., 2019), TROPOspheric Monitoring Instrument (TROPOMI) (De Smedt et al., 2018, 2021), Environmental Trace Gases Monitoring Instrument (EMI) (Su et al., 2022), and OMPS on the NOAA-20 satellite (Kwon et al., 2023; Nowlan et al., 2023). Furthermore, in the past few years, next-generation geostationary (GEO) satellite instruments (i.e., Geostationary Environment Monitoring Spectrometer (GEMS) (Kim et al., 2020; Kwon et al., 2019; Lee et al., 2024), Tropospheric Emissions: Monitoring of Pollution (TEMPO) (Zoogman et al., 2017), and Sentinel-4 (Ingmann et al., 2012)) have been designed to make hourly measurements of several trace gases including HCHO over East Asia, North America, and Europe and North Africa, respectively.

The OMI instrument (Levelt et al., 2006), which is onboard the Earth Observing System (EOS) Aura satellite, was launched by the National Aeronautics and Space Administration (NASA) in July 2004. The Aura satellite operates in an afternoon Sun-synchronous orbit with a local equator overpass time near 13:45 LT. The primary objectives of the OMI instrument have been to monitor the ozone layer, identify the origins of aerosols and trace gases, track their transport patterns, and study the influence of ozone and aerosols on climate change (Levelt et al., 2018). OMI's long-term data record has been a significant contributor in advancing our insight into stratospheric and tropospheric chemistry, air quality, and climate change. The OMI instrument measures various trace gases amounts in the Earth's atmosphere globally. This study focuses on the OMI HCHO product from the Smithsonian Astrophysical Observatory (SAO).

As OMI provides long-term measurements of HCHO, the OMI HCHO product has been extensively employed in atmospheric studies. This data set has been a useful resource for estimating variations and trends in atmospheric HCHO levels and identifying emission sources (Kuttippurath et al., 2022; Surl et al., 2018; Zhu, Mickley, et al., 2017), quantifying isoprene emissions (Kaiser et al., 2018; Marais et al., 2012, 2014; H. Wang et al., 2021), estimating surface HCHO concentrations to evaluate potential health risks (Zhu, Jacob, et al., 2017), and studying the ozone sensitivity of different regions worldwide (Duncan et al., 2010; Jin & Holloway, 2015; Jin et al., 2017; D. Li et al., 2021; Souri et al., 2017).

Previous versions of SAO OMI HCHO retrieval algorithms have been discussed by González Abad et al. (2015) and Kurosu et al. (2004). The current NASA operational OMI HCHO product using OMI Collection 3 radiances (González Abad et al., 2015) was developed and is maintained by the SAO. Within the framework of the Making Earth System Data Records for Use in Research Environments (MEaSUREs) program, this study generates the OMI Collection 4 HCHO using an updated SAO retrieval algorithm. The OMI Collection 4 HCHO uses updated spectroscopic information and updated inputs to the radiative transfer model, as well as the recently released OMI Collection 4 Level-1B radiances, which provide a long-term detrended data record of radiances (Kleipool et al., 2022). This new release is an effort to improve product stability over time, and improve the precision and accuracy of the OMI HCHO product. In addition, the OMI Collection 4 HCHO data product features

characterization of the air mass factor uncertainties, which have not been included in previous versions of the HCHO data products. The new OMI HCHO product uses the same SAO MEaSUREs algorithm as OMPS Suomi NPP and OMPS NOAA-20 HCHO products and as is shown later in this paper, their consistency and agreement are remarkable.

This paper provides a comprehensive overview of the OMI Collection 4 HCHO data product. Section 2 presents a detailed description of the data used in this study. Section 3 outlines the steps involved in the OMI Collection 4 HCHO retrieval algorithm, which includes the spectral fitting, air mass factor derivation, and reference sector correction. Additionally, Section 3 covers the uncertainty estimation and a comparison analysis between Collection 3 and Collection 4 data products. Section 4 describes validation studies of the new product using Fourier-Transform Infrared (FTIR) ground-based HCHO observations. Section 5 discusses the comparison of the OMI Collection 4 HCHO to OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI HCHO products. Finally, Section 6 summarizes the key findings and conclusions.

2. Data

2.1. OMI Observations

The NASA EOS Aura satellite was launched on 15 July 2004, into a Sun-synchronous polar orbit passing over the Equator near 13:45 local time in the ascending node. Over the first 19 years of its operational life, Aura maintained an altitude of 705 km and completed a full orbit in approximately 99 min. As a result, it conducted 14 to 15 orbits per day. Aura's orbit has started to drift since early 2023, with its orbit slowly changing (NASA Aura, 2024). The algorithm presented here will continue to work despite orbit changes. By August 2025, Aura is expected to drift by ∼50 min. Measurements remain accurate, but retrievals may be less precise due to increased solar zenith angles (NASA NSPIRES, 2022).

Onboard the Aura satellite, the OMI instrument (Levelt et al., 2006, 2018) is a nadir-looking ultraviolet-visible (UV-Vis) spectrometer that operates in a push broom configuration. It measures the Earth's backscattered radiance by employing two two-dimensional charge-coupled device (CCD) detectors from 270 to 500 nm in three channels: UV-1 (270–314 nm), UV-2 (306–380 nm), and Vis (350–500 nm) with spectral resolutions of approximately 0.42, 0.45, and 0.63 nm, respectively. The UV-2 channel is used for the HCHO retrieval. OMI has a wide across-track viewing angle of 114°, and a wide swath of 2,600 km comprising 60 across-track positions. Each orbit comprises approximately 1,650 along-track positions. In the nominal global operation mode, the ground pixel size of OMI is 13×24 km^2 near the nadir; however, moving toward the edge of the swath, the ground pixel size increases to $\sim 28 \times 150$ km^2 (González Abad et al., 2015; Levelt et al., 2006).

The OMI instrument is designed to monitor several trace gases in the Earth's atmosphere (e.g., O_3, NO_2, SO_2, HCHO, BrO, and OClO) which play crucial roles in air quality and ozone chemistry, and also collect data used to retrieve aerosols, clouds, and UV irradiance at the surface with daily global coverage (prior to the instrument row anomaly). The row anomaly in OMI refers to a radiance quality issue affecting all wavelengths for a specific viewing direction (Levelt et al., 2018; Schenkeveld et al., 2017). It is believed to be caused by factors external to the spectrometer, likely resulting from damage to the insulation blankets that cover OMI and may obstruct part of the instrument's viewing field. Early indications of the row anomaly started in 2007; however, the main impacts became apparent in 2009.

In this study, we use the OMI Collection 4 Level-1B data set. Compared to Collection 3 Level-1B, this data set has been improved using a recently developed data processor that improves the characterization of bad pixels in the spectral dimension through pixel quality flags. The Level-1B data employed in this study includes OMI Level-1B Geolocated Earthshine UV Radiance Global-mode product (OML1BRUG—version 004; Kleipool, 2021b), OMI Level-1B UV Averaged Solar Irradiances product (OML1BIRR—version 004; Kleipool, 2021a), and OMI Level-2 Cloud product (OMCLDO2—version 004; Veefkind et al., 2016).

2.2. Data for Validation and Intercomparison

2.2.1. FTIR Observations

The ground-based FTIR stations are dedicated to recording extended, inter-calibrated spectra to enhance our understanding of tropospheric and stratospheric chemistry. Regular solar absorption measurements are taken by

these stations under conditions of clear skies. Many of these stations participate in the Network for the Detection of Atmospheric Composition Change (NDACC). The NDACC FTIR HCHO retrievals (Vigouroux et al., 2018), which are used in this study, are standardized through the consistent use of spectroscopic parameters and fitting window. We assess the reliability of the OMI Collection 4 HCHO product by comparing it to ground-based FTIR HCHO measurements from 27 stations. The validation process follows the approach used by Kwon et al. (2023) for validating the OMPS Suomi NPP and OMPS NOAA-20 HCHO products. The FTIR data is available at https://www-air.larc.nasa.gov/missions/ndacc/data.html or under request to the FTIR principal investigators. More information about each station and contact information for the principal investigators can be found at https://www2.acom.ucar.edu/irwg.

2.2.2. OMPS Suomi NPP and OMPS NOAA-20 Observations

Two OMPS instruments, one launched in October 2011 aboard the Suomi NPP satellite and the other launched in November 2017 aboard the NOAA-20 satellite, fly in low Earth polar orbits with a local afternoon crossing time of 13:30 LT. The NOAA-20 satellite operates on an orbit that is 50 min behind Suomi NPP. The OMPS Suomi NPP nadir mapper measures HCHO column amounts at a spatial resolution of 50×50 km^2 at nadir. The OMPS NOAA-20 features a higher spatial resolution of 17×17 km^2 before February 2019 and 12×17 km^2 afterward, but with a lower signal-to-noise ratio than that of OMPS Suomi NPP. OMPS Suomi NPP and OMPS NOAA-20 HCHO data products are produced by SAO using the same retrieval pipeline (Nowlan et al., 2023) and inputs as the OMI HCHO presented here. This retrieval pipeline was developed as part of the MEaSUREs project to produce consistent HCHO retrievals from multiple satellite instruments. The OMPS HCHO data products are available at the NASA GES DISC (González Abad, 2022a, 2022b).

2.2.3. TROPOMI Observations

The TROPOMI instrument, launched in October 2017, onboard the low Earth orbit Sentinel-5P satellite, operates in a sun-synchronous polar orbit with a local afternoon crossing time of 13:30 LT. Using backscattered radiance in the ultraviolet, visible, near-infrared, and shortwave infrared, TROPOMI measures various trace gases and aerosols, including HCHO. The TROPOMI HCHO data product (De Smedt et al., 2018, 2021) is available with a spatial resolution of 3.5×7 km^2 from May 2018 to August 2019, and from that point onward the resolution is 3.5×5.5 km^2. For intercomparison with OMI Collection 4 HCHO, we use the TROPOMI HCHO offline data product (ESA & DLR, 2019a, 2019b), produced by the S5P Version 1 processor. The TROPOMI HCHO data product is retrieved by the Royal Belgian Institute for Space Aeronomy (BIRA-IASB).

3. The OMI Collection 4 HCHO Product

This study aims to upgrade the current NASA operational OMI HCHO product from the OMI Collection 3 retrieval to the Collection 4 version, implementing an improved retrieval algorithm. Differences between Collection 3 and Collection 4 OMI HCHO originate from improvements in the Level-1B radiance data as well as updates to the spectral fitting, air mass factor calculation, and reference sector correction steps. Table 1 presents a comprehensive summary of the differences in the spectral fitting inputs, while Table 2 provides a detailed comparison of the inputs used for air mass factor derivation in Collection 3 and Collection 4. The following section describes the retrieval algorithm.

3.1. The SAO MEaSUREs OMI HCHO Retrieval Algorithm

To retrieve OMI vertical column amounts, we follow a three-step approach: (a) differential slant column density (ΔSCD) calculation, (b) air mass factor (AMF) derivation, and (c) reference sector correction (which implements a reference background correction SCD_R, and a bias correction SCD_B) and vertical column density (VCD) calculation (Equation 1). The resulting VCD is determined by

$$VCD = \frac{\Delta SCD + SCD_R + SCD_B}{AMF}. \quad (1)$$

Furthermore, a post-processing procedure is implemented in the algorithm to include quality flags for each specific ground pixel in the retrieval product. Figure 1 illustrates the components of Equation 1 which contribute

Table 1
Fitting Parameters in OMI HCHO Retrieval

Parameter	Collection 3	Collection 4
Fitting window	328.5–356.5 nm	
Radiance reference spectrum	Computed online over the remote Pacific Ocean (30°S–30°N)	
Baseline polynomial	3rd order	
Scaling polynomial	3rd order	
Instrument slit function	Hyper-parameterization of pre-flight measurements (Dirksen et al., 2006)	Super-Gaussian parameterization (Beirle et al., 2017; Nowlan et al., 2023)
Solar reference spectrum	Chance and Kurucz (2010)	
HCHO cross sections	Chance and Orphal (2011), 300 K	
O_3 cross sections	Malicet et al. (1995), 228 and 295 K	Serdyuchenko et al. (2014), 223 and 243 K
NO_2 cross sections	Vandaele et al. (1998), 220 K	
BrO cross sections	Wilmouth et al. (1999), 228 K	
O_2–O_2 cross sections	Thalman and Volkamer (2013), 293 K	Finkenzeller and Volkamer (2022), 293 K
Molecular ring cross sections	Chance and Spurr (1997)	
Undersampling correction	Computed online (Chance et al., 2005)	
Residual common mode spectrum	Computed online (30°S–30°N)	Not included

Table 2
Inputs to AMF Calculations in OMI HCHO Retrieval

Parameter	Collection 3	Collection 4
Calculation method	Look-up table created with	Online radiative transfer calculation with
Radiative transfer model	VLIDORT V2.4 (Spurr, 2006)	VLIDORT V2.8 (Spurr, 2008)
Calculation wavelength	340 nm	
Trace gas profile	GEOS-Chem 2007 monthly climatology (calculated at 13:30 LT, 2° × 2.5°)	GEOS-Chem 2018 monthly climatology (interpolated from hourly data to overpass time, 0.5° × 0.5°)
Temperature profile	GEOS-Chem 2007 monthly climatology (calculated at 13:30 LT, 2° × 2.5°)	GEOS-Chem 2018 monthly climatology (interpolated from hourly data to overpass time, 0.5° × 0.5°)
Digital elevation model	OMI Level-1B product (OML1BRUG)	GLOBE (Hastings et al., 1999)
Surface pressure	calculated from surface altitude using NOAA (1976)	MERRA-2 (GMAO, 2015)
Vertical layer count	47-layer (GEOS-5 reduced vertical grid)	
Surface reflectance (land)	OMI surface reflectance climatology version 3 LER (Kleipool et al., 2008)	MODIS BRDF MCD43C1 product (Schaaf & Wang, 2015)
Surface reflectance (water)	OMI surface reflectance climatology version 3 LER (Kleipool et al., 2008)	Cox-Munk slope distribution (Cox & Munk, 1954)
Wind vector	Not included	MERRA-2 (GMAO, 2015)
Ocean salinity	Not included	World Ocean Atlas 2009 (Antonov et al., 2010)
Chlorophyll	Not included	MODIS Terra monthly climatology (Hu et al., 2012)
Cloud fraction	OMI Collection 3 cloud data product OMCLDO2 (Veefkind et al., 2016)	OMI Collection 4 cloud data product OMCLDO2 (Veefkind et al., 2016)
Cloud pressure	OMI Collection 3 cloud data product OMCLDO2 (Veefkind et al., 2016)	OMI Collection 4 cloud data product OMCLDO2 (Veefkind et al., 2016)
Aerosols	Not included explicitly	

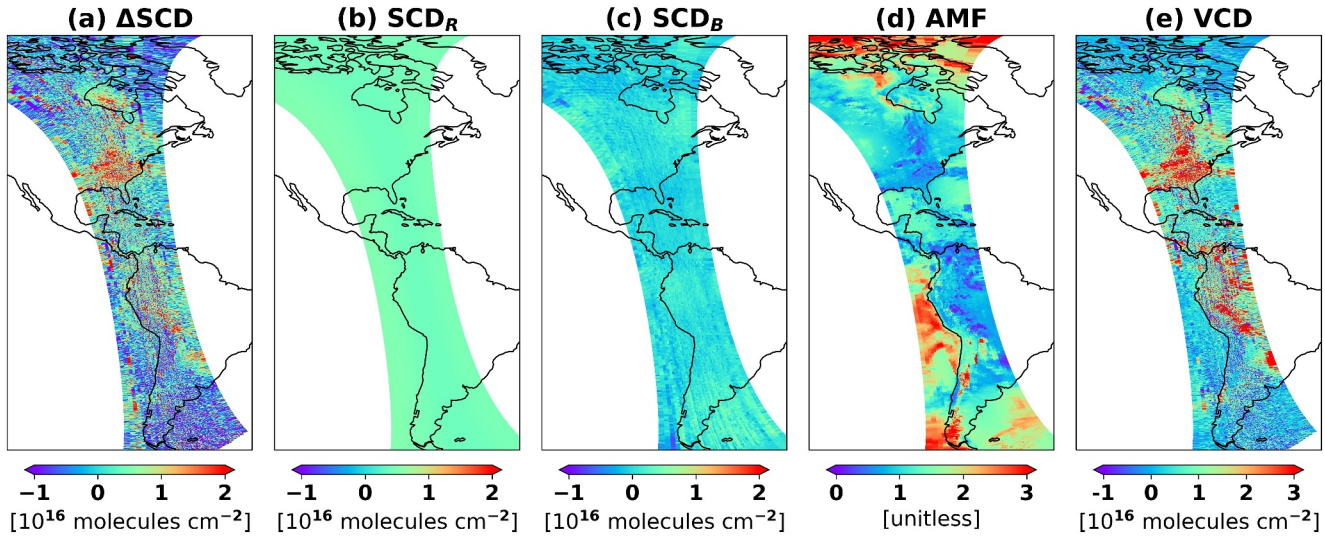

Figure 1. Components of Equation 1 for orbit 10623 observed on 14 July 2006 (all data is shown in this figure without filtering any pixels), (a) differential slant column densities (ΔSCD), (b) reference background corrections (SCD$_R$), (c) bias corrections (SCD$_B$), (d) air mass factors (AMF) and (e) vertical column densities (VCD).

to the OMI Collection 4 HCHO vertical column calculation (i.e., ΔSCD, SCD$_R$, SCD$_B$, and AMF) for orbit 10623 on 14 July 2006. A description of all three steps used to produce the VCD is provided in Sections 3.1.1–3.1.3. The algorithm is described in further detail by Nowlan et al. (2023) for its application to the OMPS instruments.

3.1.1. Spectral Fitting

The framework underlying our spectral fitting approach has been thoroughly described by Nowlan et al. (2023). In this section, we provide a summary of the spectral fitting algorithm used in the OMI Collection 4 HCHO product. We first perform a solar wavelength calibration to retrieve the OMI slit function and wavelength shift for each cross-track position. Our algorithm derives the wavelength shift by calibrating the solar irradiance measured by the OMI instrument versus a high resolution solar reference spectrum (Chance & Kurucz, 2010). We also simultaneously derive symmetric super-Gaussian terms to model the instrument slit function by convolving the high-resolution reference solar spectrum with the instrument slit function (Bak et al., 2017; Beirle et al., 2017; Nowlan et al., 2023; Sun et al., 2017).

Our algorithm then retrieves the differential slant column density of HCHO at each ground pixel using direct fitting of the modeled spectrum to the OMI measured radiance spectrum leveraging the non-linear least squares Levenberg-Marquardt minimization method following Nowlan et al. (2023) and González Abad et al. (2015). We build the modeled radiance spectrum based on a radiance reference spectrum instead of solar irradiance measurements to avoid cross-track dependent striping in the HCHO retrieval. On a daily basis, at each cross-track position, we generate the radiance reference spectrum by calculating the average of all radiance spectra sharing the same cross-track position measured over a clean area of background HCHO in the remote Pacific Ocean (30°S < latitudes < 30°N and 140°W < longitudes < 180°W).

In formulating the modeled radiance spectrum equation, we consider atmospheric scattering, surface reflectivity, and broadband instrument calibration artifacts through the baseline and scaling polynomial coefficients. We further incorporate trace gas absorption, which includes HCHO, O_3, NO_2, BrO, and O_2-O_2, using the Beer-Lambert law. We also integrate correction for spectral undersampling (Chance et al., 2005) and pre-computed Ring spectrum (Chance & Spurr, 1997). Additionally, to mitigate the effects of the South Atlantic anomaly, our algorithm applies an iterative approach to exclude pixels with unreasonably high fitting residuals (González Abad et al., 2015; Richter et al., 2011). Following the initial spectral fitting, we remove spectral pixels with residuals greater than 3 times the standard deviation of the mean fitting residuals and repeat the spectral fitting of the remaining spectral points for a maximum of four times.

The HCHO retrieval is performed using a fitting window of 328.5–356.5 nm which is selected with the aim to reduce the impact of dominant O_3 absorption at lower wavelengths and to minimize the correlation of HCHO with

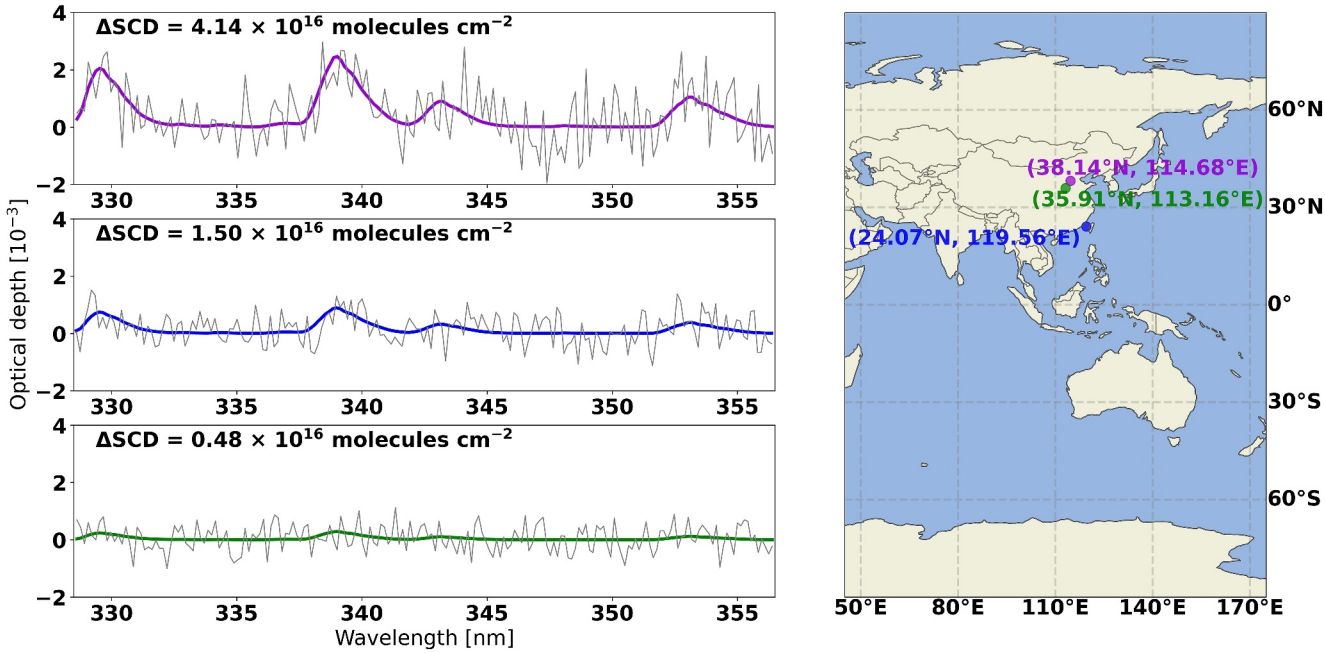

Figure 2. Differential slant optical depths of fitted HCHO for orbit 10615 on 14 July 2006 for high (top), medium (middle), and low (bottom) HCHO amounts. The colored lines show the modeled optical depths and the gray lines are the measured optical depths.

other fitting molecules, particularly BrO (González Abad et al., 2015). Table 1 provides the characteristics and inputs of the spectral fitting algorithm used in the OMI Collection 4 HCHO along with their corresponding values used in the OMI Collection 3 HCHO product. Figure 2 shows the modeled and measured optical depths of HCHO for three ground pixels, whose exact locations are marked on the right panel, in orbit 10615 on 14 July 2006. This figure compares the fitting signal and residuals at those three pixels, each representing different levels of HCHO amounts: high ($\Delta SCD = 4.14 \times 10^{16}$ molecules cm^{-2}, top left panel), medium ($\Delta SCD = 1.50 \times 10^{16}$ molecules cm^{-2}, middle left panel), and low ($\Delta SCD = 0.48 \times 10^{16}$ molecules cm^{-2}, bottom left panel). The purpose is to demonstrate the ability of the fitting algorithm to detect the HCHO signal over polluted pixels. The bottom panel shows how the HCHO signal to noise ratio decreases in less polluted areas.

3.1.2. Air Mass Factor Calculation

The air mass factor accounts for the mean photon path through the atmosphere, enabling the conversion of slant column density to vertical column density. In general terms, the VCD is derived by $VCD = SCD/AMF$. The AMF derivation is explained in detail by Nowlan et al. (2023) and we follow the same steps. The OMI Collection 4 HCHO retrieval performs an online AMF calculation for each ground pixel following the method presented by Palmer et al. (2001). Equation 2 summarizes this approach which can be used for optically thin absorbers including HCHO:

$$\mathrm{AMF} = \int_z w(z)s(z)dz. \tag{2}$$

The AMF calculation involves determining two key components: scattering weights ($w(z)$) and shape factors ($s(z)$). Scattering weights define the sensitivity of satellite observations to atmospheric layers at various altitudes (z), while shape factors characterize the vertical distribution of trace gases. The shape factor component at each layer z is defined as the partial column of HCHO at that layer normalized by the total column of HCHO. In order to calculate the AMFs for a partially cloudy pixel, we use the independent pixel approximation, presented by Martin et al. (2002).

Table 2 summarizes the inputs provided to the AMF calculations in the OMI Collection 4 HCHO retrieval algorithm, along with corresponding inputs to the OMI Collection 3 HCHO AMF algorithm. The retrieval

algorithm in Collection 4 performs an online scene by scene calculation to retrieve the AMF at each ground pixel, while the Collection 3 algorithm (González Abad et al., 2015) relied on pre-determined look-up tables to calculate AMFs. In what follows in this section, we discuss the inputs used by Collection 4 in the radiative transfer model and compare them with inputs used by Collection 3.

We determine shape factors using HCHO and temperature vertical profiles generated by the GEOS-Chem chemical transport model (Bey et al., 2001). The OMI Collection 4 algorithm uses the GEOS-Chem simulation, driven by meteorological fields from the Modern-Era Retrospective analysis for Research and Applications Version 2 (MERRA-2; Gelaro et al., 2017), to construct $0.5° \times 0.5°$ monthly mean climatology of hourly HCHO and temperature profiles in the year 2018. Those profiles are then used to interpolate HCHO and temperature profiles at the satellite overpass time. This is an improvement compared to the HCHO and temperature profiles used in the OMI HCHO Collection 3 AMF calculations (González Abad et al., 2015) which featured profiles with $2° \times 2.5°$ spatial resolution calculated at 13:30 LT by GEOS-Chem 2007 simulation. Additionally, we use the same GEOS-Chem 2018 model to produce corresponding monthly mean climatologies of hourly ozone profiles which are used in the Collection 4 radiative transfer calculation, while Collection 3 used fixed ozone profiles selected based on latitude and date.

We use the Vector Linearized Discrete Ordinate Radiative Transfer (VLIDORT) model Version 2.8 (Spurr, 2008) to determine scattering weights at 340 nm on the GEOS-5 (Goddard Earth Observing System Model, Version 5) reduced grid of 47 vertical layers ranging from the surface to 0.01 hPa (GEOS-Chem Support Team, 2018). The radiative transfer code uses inputs of surface pressure, surface reflectance, ozone concentration, cloud fraction, and cloud pressure. In Collection 4, we adjust the hourly surface pressures from MERRA-2 corresponding to each satellite overpass time using a terrain height adjustment (Boersma et al., 2011; Zhou et al., 2009) to correct for differences in spatial resolution with the satellite ground pixels. We use the Global Land One-kilometer Base Elevation (GLOBE) digital elevation model (Hastings et al., 1999) from NOAA for this correction. The partial column profiles used in the AMF calculation are also updated accordingly. This is a departure from the radiative transfer calculation in Collection 3 which relied on the OMI Level-1B product (OML1BRUG) supplied terrain height to derive surface pressure using the US Standard Atmosphere from NOAA (1976).

The OMI Collection 3 algorithm determined surface reflectance by interpolating the Lambertian equivalent reflectance (LER) field from the OMI surface reflectance climatology version 3 (Kleipool et al., 2008) at the wavelength of 340 nm, which is the same wavelength used for calculating AMFs. However, in Collection 4, depending on whether the scene's surface is land or water, we use different approaches to derive the surface reflectance (Chong et al., 2024; Nowlan et al., 2023; H. Wang et al., 2023). Over land, we use the bidirectional reflectance distribution function (BRDF) data product from the MODerate Resolution Imaging Spectroradiometer (MODIS) (MCD43C1; Schaaf & Wang, 2015) and extend it to 340 nm as explained in Nowlan et al. (2023).

The Cox-Munk slope distribution determines the surface reflectance over water (Cox & Munk, 1954). This is derived from the VBRDF package in VLIDORT (Spurr & Christi, 2019), utilizing wind speed and ocean salinity. MERRA-2 provides $0.5° \times 0.625°$ hourly wind properties at a height of 2-m (GMAO, 2015), and the World Ocean Atlas 2009 provides $0.1° \times 0.1°$ monthly mean climatologies of ocean salinity (Antonov et al., 2010). Additionally, the VSLEAVE package in VLIDORT (Spurr & Christi, 2019) calculates surface leaving radiances from water, using inputs such as chlorophyll concentration, ocean salinity, observation geometries, and wind speed. We obtain the monthly mean climatology of chlorophyll from MODIS Terra 18-year (2000–2018) with a spatial resolution of 9.28 km (Hu et al., 2012).

Our radiative transfer code uses MODIS BRDF data to account for snow and ice surface reflectance (Howlett et al., 2023). The OMI Collection 4 HCHO Level-2 product provides snow and ice fractions in the Level-2 data files for evaluation purposes. We determine the snow and sea ice fractions in the northern hemisphere for each ground pixel using the Interactive Multisensor Snow and Ice Mapping System (IMS) data product with 4 km resolution developed by the U.S. National Ice Center (2008). For the southern hemisphere, we use the percent snow cover data from the MODIS MCD43C1 product for snow fraction, and calculate sea ice fraction from the daily Sea Ice Index product with a 25 km × 25 km spatial resolution from the National Snow and Ice Data Center (NSIDC) (Fetterer et al., 2017).

Table 3
Pixel Quality Flags Criteria

Pixel quality flag	Description
Bad (main data quality flag = 2)	$\|VCD\| > 2 \times 10^{17}$ molecules cm^{-2} or $VCD + 3\sigma_{VCD} < 0$ or $AMF < 0.1$ or geometric $AMF > 5$
Suspect (main data quality flag = 1)	$VCD + 2\sigma_{VCD} < 0$ or geometric $AMF > 4$
Good (main data quality flag = 0)	$\|VCD\| \leq 2 \times 10^{17}$ molecules cm^{-2} and $VCD + 2\sigma_{VCD} \geq 0$ and $AMF \geq 0.1$ and geometric $AMF \leq 4$

Cloud fractions and cloud pressures for radiative transfer calculations are from the OMI Collection 4 cloud product (OMCLDO2; Veefkind et al., 2016). When performing the radiative transfer calculation, we treat clouds as Lambertian surfaces assuming they have a constant albedo of 0.8. In addition, our algorithm implicitly considers the presence of scattering aerosols through the use of the cloud product (Nowlan et al., 2023; H. Wang et al., 2023). However, since the AMF simulation of the OMI Collection 4 HCHO product does not explicitly include aerosols, this may introduce uncertainty in the HCHO column amount in heavily aerosol-polluted areas (Jung et al., 2019; Kwon et al., 2017).

3.1.3. Reference Sector Correction

As described by Nowlan et al. (2023) and Chong et al. (2024), we perform the reference sector correction in two parts: (a) correction for background HCHO, and (b) bias correction. We derive an HCHO background correction for each orbit in order to compensate for the background HCHO SCD (SCD_R) present in the radiance reference spectrum. Over the reference sector region, we determine the modeled VCDs using the GEOS-Chem monthly climatology vertical profiles and then we calculate the modeled SCDs by using the AMFs for each ground pixel. Then, we determine the median value of those SCDs at each cross-track position and apply a third-order polynomial fitting to obtain a smoothed SCD_R as a function of cross-track. This smoothed SCD_R is then applied to each ground pixel based on cross-track position.

In addition, we also calculate a bias correction for each orbit (Nowlan et al., 2023). This correction aims to mitigate biases that may arise due to factors like stronger ozone absorption, unaccounted instrument calibration, or retrieval physics. In the case of OMI HCHO, the bias correction is minimal. However, it is included for consistency with retrievals from other instruments that require larger latitude-dependent bias correction (Nowlan et al., 2023). Our algorithm selects 30 orbits over the Pacific closest in date to the target orbit. For each 1° latitudinal and 1° solar zenith angle bin, we calculate the mean difference between corrected SCDs ($\Delta SCD + SCD_R$) and modeled SCDs, excluding outliers, and perform wavelet denoising to derive the bias correction (SCD_B) which is applied to each ground pixel based on latitude and solar zenith angle. We detect outliers using a window of 30 cross-track and 15 along-track positions, considering a data point as an outlier if it deviates by more than 3 times the standard deviation from the window's median.

The final stage of our algorithm evaluates the quality of retrieval at each ground pixel by assigning quality flags of 0 (good), 1 (suspect), or 2 (bad). We define pixel quality flags following the logic provided in Table 3. In this table, $\sigma_{V\ CD}$ is the error in VCD associated with the fitting error in the SCD. Additionally, as a general guideline, we advise against using the OMI Collection 4 HCHO product in situations with high cloud fraction (cloud fraction > 0.4) or high solar zenith angles (SZA > 70°) due to the likelihood of significant biases. In addition, we recommend exercising extreme caution when using pixels covered with snow or ice as there may be large uncertainties in the cloud products.

Figure 3 displays the global HCHO differential slant column densities, air mass factors, and vertical column densities for Collection 4, averaged over June 2006 and regridded to 0.1° × 0.1° resolution applying the physical oversampling approach (Sun et al., 2018). To ensure data reliability, pixels with solar zenith angles above 70° and cloud fractions greater than 0.4, and main data quality flags marked as bad are excluded from the data used to generate the plot. Regional enhancement of HCHO concentrations are clearly visible from biomass burning in Africa, anthropogenic sources in northern India and eastern China, and isoprene emissions in the southeastern US. Enhanced columns in northern high latitudes result from wildfires in Siberia, Alaska, and northern Canada (Zhao et al., 2024). The blue boxes in Figure 3 represent the regions used later in Sections 3.3 and 5 for comparing OMI

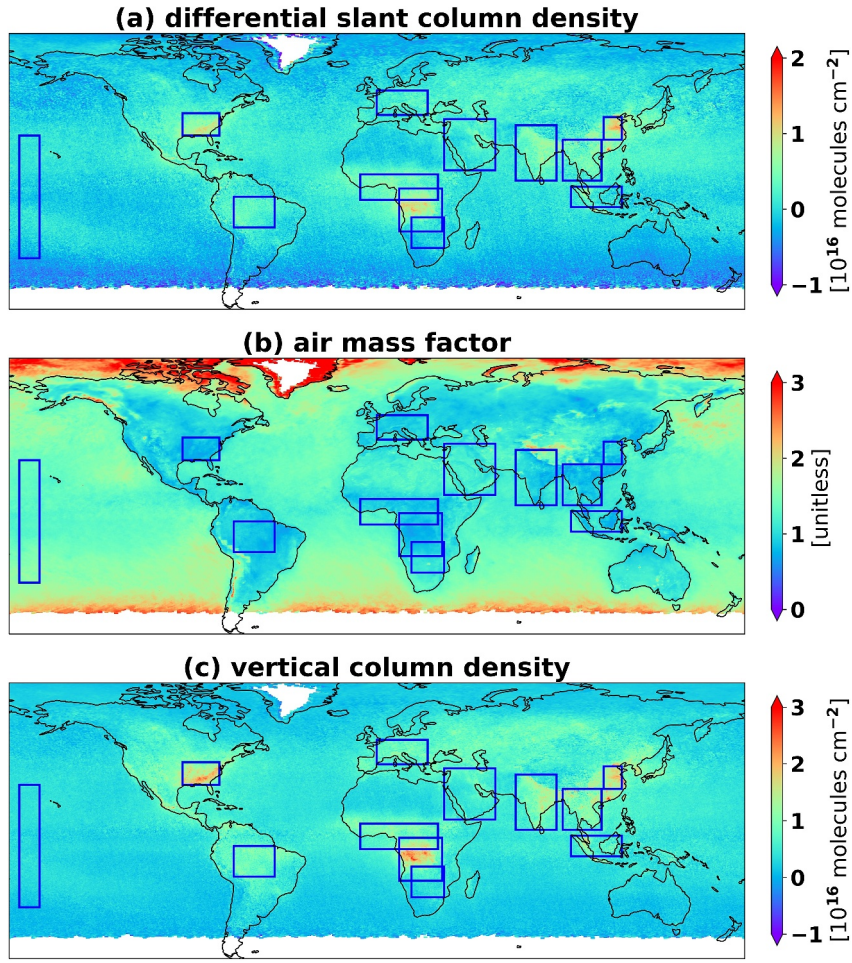

Figure 3. June 2006 mean (a) HCHO differential slant column densities, (b) Air mass factors, and (c) Vertical column densities. Mean values are calculated for pixels with solar zenith angle < 70° and cloud fractions < 0.4, regridded to 0.1° × 0.1° resolution. The blue boxes illustrate the regions used later in Figures 6, 7, 14, and 15 and Table 7. The geographic boundaries of these boxes are defined in Table 4.

Collection 4 HCHO with OMI Collection 3 HCHO and for intercomparisons with TROPOMI and OMPS HCHO, respectively. Table 4 provides the geographic boundaries of these boxes.

3.2. Uncertainties

We assess the random (precision) and systematic (bias) uncertainties associated with the OMI Collection 4 HCHO data product in this section. Previous versions of the OMI HCHO products only provided estimates of the slant column random uncertainty. We present here an effort to expand the uncertainty budget beyond evaluations in previous studies to estimate uncertainties associated with the air mass factors on a pixel basis and to include them in the product files. We discuss the different sources of random and systematic uncertainties in the following subsections.

3.2.1. Slant Column Density Uncertainties

The primary contributor to random fitting uncertainties associated with HCHO differential slant column densities is the instrument measurement noise. For the OMI Collection 4 HCHO data product, within pixels characterized by solar zenith angles less than 70° and cloud fractions smaller than 0.4, the typical fitting uncertainty in differential slant column density is around $\sim 6.5 \times 10^{15}$ molecules cm^{-2}. Figures 4a and 4d illustrate histograms of random uncertainties in HCHO differential slant column densities retrieved from OMI observations in January

Table 4
Geographic Boundaries for Regions Illustrated in Figure 3

Region	Latitude boundaries	Longitudes boundaries
Pacific Ocean	30°S–30°N	175°W–165°W
Southeast US	30°N–41°N	95°W–77°W
Amazon Basin	15°S–0°	70°W–50°W
Europe	40°N–52°N	0°–25°E
Middle East	13°N–38°N	33°E–58°E
India	8°N–35°N	68°E–88°E
West-Central Africa	1.5°S–11°N	8°W–30°E
Central Africa	17°S–4°N	11°E–32°E
Southern Africa	25°S–10°S	17°E–33°E
East China	28°N–39°N	111°E–120°E
Southeast Asia	8°N–28°N	91°E–110°E
Equatorial Asia	5°S–5°N	95°E–120°E

and July 2006, respectively. In January (July) 2006, the HCHO differential SCDs have a median random uncertainty of 6.12×10^{15} molecules cm^{-2} (6.30×10^{15} molecules cm^{-2}).

The uncertainties associated with the reference background correction and bias correction affect corrected HCHO slant column densities. Random uncertainties in the reference background correction originate from the uncertainties in the modeled HCHO column densities and corresponding AMFs used to obtain modeled SCDs. Errors in AMFs are estimated in Section 3.2.2. As discussed by Nowlan et al. (2023), we assume that uncertainty in the modeled HCHO column amounts is negligible. However, we account for the natural variability in the modeled HCHO columns within the reference sector, using the median absolute deviation (MAD) to measure this variability. To calculate the random uncertainties in the reference sector correction, we combine the natural variability of the modeled HCHO columns with the uncertainty in the AMFs using Gaussian error propagation.

Random uncertainties in the bias correction arise from modeled HCHO column densities, corresponding AMFs, and also the uncertainties associated with corrected SCDs (ΔSCD + SCD$_R$). As uncertainty in the modeled HCHO VCDs is negligible and uncertainty in SCD$_R$ has been already accounted for, to estimate the random uncertainties in the bias correction, we incorporate the uncertainties from the AMFs and ΔSCDs, employing Gaussian error propagation.

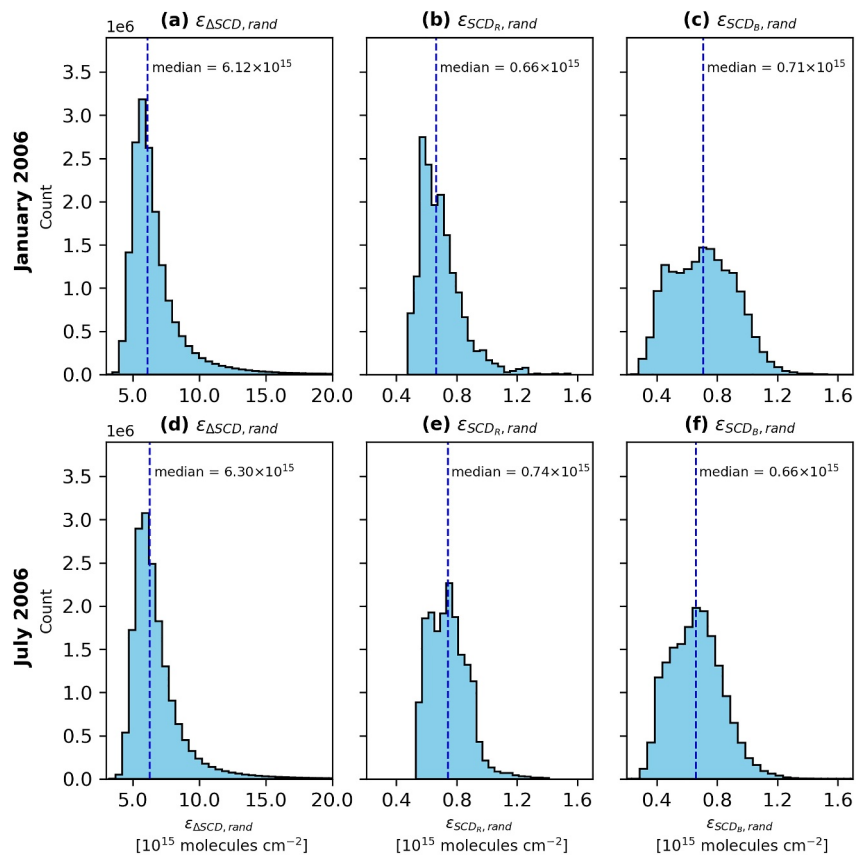

Figure 4. Histograms of random uncertainties in (a), (d) differential slant column density, (b, e) reference background correction, and (c, f) bias correction for pixels with solar zenith angle <70° and cloud fractions < 0.4 in January 2006 (top panels) and July 2006 (bottom panels). The median of the random uncertainties is printed in the plots for each component. The blue dashed line shows the median of uncertainty at each panel.

Figures 4b and 4e display the reference background correction random uncertainties for OMI HCHO data in January and July 2006 as histograms. As shown in these figures, the median random uncertainty in reference background correction is 0.66×10^{15} molecules cm^{-2} (0.74×10^{15} molecules cm^{-2}) in January (July) 2006. Similarly, the random uncertainties in bias correction for the same OMI pixels are displayed in Figures 4c and 4f. The median random uncertainty in bias correction is 0.71×10^{15} molecules cm^{-2} (0.66×10^{15} molecules cm^{-2}) in January (July) 2006.

The OMI Collection 4 HCHO slant column density random uncertainty appears to show low degradation over time. Compared to the uncertainty values reported for January (July) 2006 in the previous paragraphs, 10 years later in January (July) 2016 median random uncertainties in differential slant column densities, reference background correction, and bias correction are 6.44×10^{15} molecules cm^{-2} (6.71×10^{15} molecules cm^{-2}), 0.72×10^{15} molecules cm^{-2} (0.75×10^{15} molecules cm^{-2}), and 0.66×10^{15} molecules cm^{-2} (0.63×10^{15} molecules cm^{-2}), respectively.

In addition to random uncertainties, there are systematic uncertainties associated with the HCHO slant column densities. There are several contributors to the HCHO slant column density systematic uncertainty. These include uncertainties associated with errors in the slit function, absorption cross sections, wavelength calibration, as well as the selections of wavelength fitting window and polynomial fitting order (Nowlan et al., 2023). Additionally, there are systematic errors due to the reference background correction and bias correction. De Smedt et al. (2018) suggest that the systematic uncertainty in the OMI HCHO slant column densities measured from 2005 to 2014 is estimated to be 25% over remote clean regions and 15% over regions with enhancements of HCHO.

3.2.2. Air Mass Factor Uncertainties

In this section, we provide estimations of the air mass factor uncertainties. The estimation of AMF uncertainties is a complex process. This is because it can be influenced by a range of errors stemming from multiple input parameters. However, leveraging experience from previous studies, we develop an approach to quantify random and systematic uncertainties linked to air mass factors for every OMI observation pixel. Chong et al. (2024) used a similar approach to estimate AMF random uncertainties in the SAO OMPS BrO retrieval.

Equation 3 calculates the random uncertainties associated with air mass factors ($\varepsilon_{AMF,rand}$) assuming no correlations between different components contributing to the AMF random uncertainties. These error sources include the HCHO shape factor ($\varepsilon_{sf,rand}$), surface reflectance ($\varepsilon_{sr,rand}$), cloud fraction ($\varepsilon_{cf,rand}$) and cloud pressure ($\varepsilon_{cp,rand}$). In Equation 3, $\varepsilon_{AMF,rand}$ represents the air mass factor random uncertainty at each OMI pixel. The indices sf, sr, cf, and cp represent the shape factor, surface reflectance, cloud fraction, and cloud pressure, respectively.

$$\varepsilon^2_{AMF,rand} = \varepsilon^2_{sf,rand} + \left(\frac{\partial AMF}{\partial sr}\right)^2 \varepsilon^2_{sr,rand} + \left(\frac{\partial AMF}{\partial cf}\right)^2 \varepsilon^2_{cf,rand} + \left(\frac{\partial AMF}{\partial cp}\right)^2 \varepsilon^2_{cp,rand} \quad (3)$$

As the first step, we determine the random uncertainty associated with the HCHO shape factor error. As discussed in detail by Chong et al. (2024), we follow an empirical approach to categorize HCHO vertical profiles generated by GEOS-Chem into four clusters by applying the k-means clustering (Lloyd, 1982). Considering the shape of the HCHO vertical profile at each pixel, we categorize the profile into one of these four profile clusters and assign a cluster index (1, 2, 3, 4) to the pixel. We then calculate the standard deviation of the AMFs assigned to each profile cluster index to provide an estimation of $\varepsilon_{sf,rand}$. This way, we provide an (over)estimation of the uncertainty associated with the natural variability of the model but not the model noise.

To estimate $\varepsilon_{sf,rand}$ and partial derivatives ($\frac{\partial AMF}{\partial sr}$, $\frac{\partial AMF}{\partial cf}$, and $\frac{\partial AMF}{\partial cp}$) in Equation 3, we leverage the entire year of 2006 OMI HCHO AMF data. This data set serves as the foundation for generating look-up tables to estimate $\varepsilon_{sf,rand}$ and partial derivatives. We classify AMFs based on five key parameters: HCHO vertical profile cluster index, geometric AMF, surface reflectance, cloud fraction, and cloud pressure. These parameters form bins with specific values (nodes and intervals) as outlined in Table 5. Additionally, we categorize AMFs according to the pixel surface type, including land, water, and glint (specular reflection angle <30°). Within this framework, we sort out AMFs from all pixels in 2006 into their respective bins. Subsequently, for each bin, we compute the standard deviation of AMFs, as an estimation of $\varepsilon_{sf,rand}$. Additionally, we calculate the mean AMF for each bin.

Table 5
Nodes and Intervals for Vertical Profile Cluster, Geometric AMF, Surface Reflectance, Cloud Fraction, and Cloud Pressure Which Are Used to Create Bins for the AMF Uncertainties Look-Up Tables

Parameter	Nodes and intervals
Vertical profile cluster	1 to 4 with 1 interval
Geometric AMF	2.0 to 5.5 with 0.5 interval
Surface reflectance	0.0 to 1.0 with 0.05 interval
Cloud fraction	0.0 to 1.0 with 0.1 interval
Cloud pressure	100 to 1,100 hPa with 150 hPa interval

To determine the partial derivatives in Equation 3, we calculate the gradients of the mean AMF with respect to surface reflectance, cloud fraction, and cloud pressure using neighboring bins. Then, we assign these gradient values, denoted as $\frac{\partial AMF}{\partial sr}$, $\frac{\partial AMF}{\partial cf}$, and $\frac{\partial AMF}{\partial cp}$, together with the standard deviations of AMFs ($\varepsilon_{sf,rand}$) to their corresponding bin in order to form four distinct look-up tables (one for each partial derivative term and one for $\varepsilon_{sf,rand}$). Ultimately, for every orbit, we employ these lookup tables to interpolate and estimate the partial derivatives ($\frac{\partial AMF}{\partial sr}$, $\frac{\partial AMF}{\partial cf}$, and $\frac{\partial AMF}{\partial cp}$) and random uncertainty in shape factors ($\varepsilon_{sf,rand}$) at each Level-2 pixel.

We determine the systematic uncertainties in air mass factor ($\varepsilon_{AMF,syst}$) using Equation 4. In this equation, $\varepsilon_{sr,syst}$, $\varepsilon_{cf,syst}$, and $\varepsilon_{cp,syst}$ represent systematic uncertainties in the surface reflectance, cloud fraction, and cloud pressure, respectively. Given that each component of systematic uncertainty can be either positive or negative, we estimate the maximum absolute value for AMF systematic uncertainty under a conservative assumption: all sources of systematic uncertainties collectively affect the AMF in a uniform direction. The shape factor systematic uncertainties are not included in this equation as our knowledge about shape factor biases is limited. However, the product includes scattering weights, allowing users to use vertical profiles from other models and compare the results.

$$\varepsilon^2_{AMF,syst} = \left(\frac{\partial AMF}{\partial sr}\right)^2 \varepsilon^2_{sr,syst} + \left(\frac{\partial AMF}{\partial cf}\right)^2 \varepsilon^2_{cf,syst} + \left(\frac{\partial AMF}{\partial cp}\right)^2 \varepsilon^2_{cp,syst} \quad (4)$$

In Equations 3 and 4, the surface reflectance uncertainties ($\varepsilon_{sr,rand}$, $\varepsilon_{sr,syst}$) are functions of surface type. For OMI pixels over land, we determine the surface type using the Land Cover Climate Modeling Grid (CMG) product form MODIS (MCD12C1 Version 6) (Friedl & Sulla-Menashe, 2019). Each land surface type is associated with specific surface reflectance random and systematic uncertainties which are taken from previous studies (Z. Wang et al., 2018; Wu et al., 2018). Additionally, for OMI pixels over water, we use the uncertainty values reported by Fasnacht et al. (2019). Table 6 summarizes the typical values, random and systematic uncertainties associated with each surface type. Typical values of surface reflectivity are estimated based on averaged MODIS BRDF band 3 isotropic values for each land type. For oceans and open water, surface reflectivity typical value is estimated according to Tilstra et al. (2017). For those MODIS surface types which lack uncertainty estimates, we assign the highest value within comparable surface categories summarized by Z. Wang et al. (2018), with the exception of Deciduous Needleleaf Forests, which adopt the same value as Evergreen Needleleaf Forests. We obtain the random and systematic uncertainties in cloud fraction ($\varepsilon_{cf,rand}$, $\varepsilon_{cf,syst}$) and cloud pressure ($\varepsilon_{cp,rand}$, $\varepsilon_{cp,syst}$) from the OMI Level-2 cloud data product (OMCLDO2) (Veefkind et al., 2016).

Figures 5a–5d provide global overviews of the estimated AMF random uncertainties (Figures 5a and 5c) and absolute values of AMF systematic uncertainties (Figures 5b and 5d) derived by Equations 3 and 4 for OMI observations on 14 January 2006 and 14 July 2006. Additionally, we collect 2 months of AMF random uncertainties and absolute values of AMF systematic uncertainties for all OMI pixels with solar zenith angles below 70° and cloud fractions smaller than 0.4 in January 2006 and July 2006 to produce Figures 5e–5h. These figures display the distributions of AMF random uncertainties (Figures 5e and 5g) and absolute values of AMF systematic uncertainties (Figures 5f and 5h) for those pixels as histograms. In January (July) 2006, the median AMF random uncertainty is 11.3% (13.0%) and the median AMF systematic uncertainty is 8.1% (8.9%). The histograms of AMF random uncertainties and systematic uncertainties reveal two distinct peaks, each corresponding to specific observation conditions. The first peak primarily reflects AMF uncertainties in pixels situated over water bodies, as well as pixels with relatively low cloud fractions (<0.1). On the other hand, the second peak is predominantly formed by pixels over land and pixels with higher cloud fractions (>0.1), resulting in higher levels of AMF uncertainties.

The median AMF random uncertainty and the median AMF systematic uncertainty are similar in later years. For example, the random uncertainty is 11.3% (13.7%) and systematic uncertainty is 9.2% (8.9%) in January (July) 2016. The estimated AMF random and systematic uncertainties associated with individual OMI observations vary due to the variability in AMF input parameters, especially cloud fractions and cloud pressure. However, overall,

Table 6
Typical Value, Random and Systematic Uncertainties of Surface Reflectance Parameter as a Function of MODIS Land Cover Type

Surface	Typical value	Random uncertainty	Systematic uncertainty
Water Bodies	0.05	0.018 (Fasnacht et al., 2019)	−0.015 (Fasnacht et al., 2019)
Evergreen Needleleaf Forests	0.07	0.0237 (Z. Wang et al., 2018)	−0.0076 (Z. Wang et al., 2018)
Evergreen Broadleaf Forests	0.02	0.0196 (Z. Wang et al., 2018)	−0.0009 (Z. Wang et al., 2018)
Deciduous Needleleaf Forests	0.12	0.0237 (maximum used)	−0.0076 (maximum used)
Deciduous Broadleaf Forests	0.05	0.0196 (Z. Wang et al., 2018)	−0.0009 (Z. Wang et al., 2018)
Mixed Forests	0.08	0.0201 (Z. Wang et al., 2018)	0.0015 (Z. Wang et al., 2018)
Closed Shrublands	0.05	0.0125 (Z. Wang et al., 2018)	0.0071 (Z. Wang et al., 2018)
Open Shrublands	0.08	0.0318 (Z. Wang et al., 2018)	0.0069 (Z. Wang et al., 2018)
Woody Savannas	0.06	0.0318 (maximum used)	−0.0076 (maximum used)
Savannas	0.04	0.0125 (Z. Wang et al., 2018)	0.0071 (Z. Wang et al., 2018)
Grasslands	0.08	0.0318 (Z. Wang et al., 2018)	0.0069 (Z. Wang et al., 2018)
Permanent Wetlands	0.05	0.0318 (maximum used)	−0.0076 (maximum used)
Croplands	0.06	0.0318 (Z. Wang et al., 2018)	0.0069 (Z. Wang et al., 2018)
Urban and Built-up Lands	0.07	0.0318 (maximum used)	−0.0076 (maximum used)
Cropland/Natural Vegetation Mosaics	0.05	0.013 (Wu et al., 2018)	0.013 (Wu et al., 2018)
Permanent Snow and Ice	0.39	0.0505 (Z. Wang et al., 2018)	−0.0449 (Z. Wang et al., 2018)
Barren	0.15	0.0111 (Z. Wang et al., 2018)	0.0023 (Z. Wang et al., 2018)

Note. When a surface type uses a maximum value for a missing uncertainty, the value is the maximum uncertainty presented by Z. Wang et al. (2018).

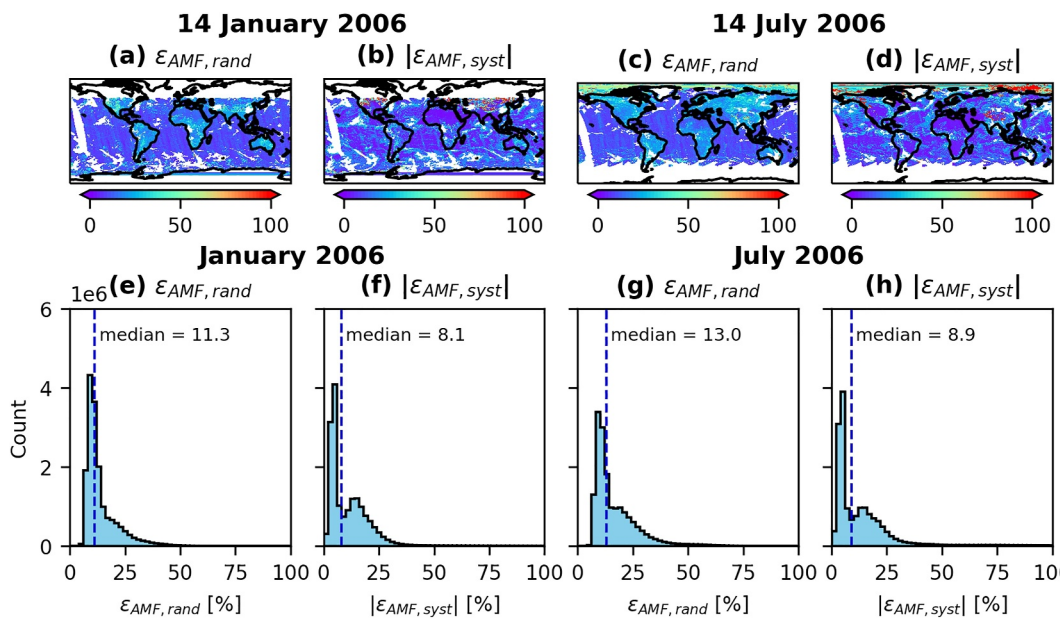

Figure 5. (a, c) AMF random uncertainties, and (b, d) absolute values for AMF systematic uncertainties associated with measurements on 14 January 2006 and 14 July 2006. (e, g) Histograms of AMF random uncertainties, and (f, h) histograms of absolute values of AMF systematic uncertainties calculated for pixels with solar zenith angle < 70° and cloud fractions < 0.4 in January 2006 and July 2006. The median of the random and systematic uncertainties is printed in the plots. The blue dashed line shows the median of uncertainty at each panel.

the medians of AMF random and systematic uncertainties remain relatively constant over time and are on the order of $\sim 11 - 14\%$, and $\sim 7 - 11\%$, respectively.

3.2.3. Vertical Column Density Uncertainties

Equation 5 calculates the random uncertainty in the final HCHO column amount product. In this equation, $\varepsilon_{\mathrm{VCD,rand}}$, $\varepsilon_{\Delta SCD,rand}$, $\varepsilon_{SCD_R,rand}$, $\varepsilon_{SCD_B,rand}$, and $\varepsilon_{AMF,rand}$ represent the random uncertainties in vertical column density, differential slant column density, reference background correction, bias correction, and air mass factor, respectively. Using Equation 5, we estimate the median random uncertainty to be on the order of $\sim 5 \times 10^{15}$ molecules cm^{-2} in the OMI Collection 4 HCHO VCD product for pixels with solar zenith angles below 70° and cloud fractions smaller than 0.4.

$$\varepsilon^2_{\mathrm{VCD,rand}} = \left(\frac{1}{\mathrm{AMF}}\right)^2 \left(\varepsilon^2_{\Delta SCD,rand} + \varepsilon^2_{SCD_R,rand} + \varepsilon^2_{SCD_B,rand}\right) + \left(\frac{\mathrm{SCD}}{\mathrm{AMF}^2}\right)^2 \varepsilon^2_{AMF,rand} \tag{5}$$

Given the potential for each component of systematic uncertainty to deviate from a Gaussian distribution and be correlated with other components, we cannot use Equation 5 to estimate the systematic uncertainties in the VCDs. To have an estimation of systematic uncertainties associated with the final products, we recommend reviewing the estimated systematic uncertainties associated with SCDs and AMFs discussed in Sections 3.2.1 and 3.2.2, respectively.

To estimate uncertainties in gridded Level 3 OMI HCHO data, standard error propagation should be used for random uncertainties. Random uncertainties in gridded data decrease as more observations are combined. Systematic uncertainties, however, can contribute in more complex ways. We recommend avoiding weighting data by uncertainties to avoid sampling biases which can induce artificial geophysical variability over time.

3.3. Comparison Between Collection 3 and Collection 4 OMI HCHO Products

The differences between OMI HCHO in Collection 3 and Collection 4 are due to improvements in the Level-1B radiance data, as well as updates to the spectral fitting (Table 1), air mass factor calculation (Table 2), and reference sector correction steps. Kleipool et al. (2022) recently released the Collection 4 Level-1B data set, which uses a newly developed data processor. The improvements in the Level-1B data set, particularly the characterization of bad pixels in the spectral dimension through pixel quality flags, have a significant impact on the Level-2 OMI HCHO product. The spectral fitting process in Collection 4 benefits from updated molecular absorption cross-section and input parameters. The radiative transfer model in Collection 4 performs scene-by-scene online calculations, as opposed to Collection 3, which relied on pre-determined look-up tables. Among other updates, the radiative transfer model uses updated inputs with higher spatial resolution. In Collection 4, the reference sector correction for OMI HCHO employs a two-step approach to account for background HCHO correction and latitudinal bias separately, as explained in Section 3.1.3, while the retrieval algorithm in Collection 3 included reference sector and latitudinal bias corrections in one single variable to address both (González Abad et al., 2015). Splitting them in Collection 4 provides flexibility to develop each step tailored to each component.

In Figures 6 and 7, we compare the stability of Collection 3 and 4 products over the period 2005 to 2019 for six regions: Pacific Ocean, Southeast US, Amazon Basin, Europe, East China, and Central Africa (see blue boxes in Figure 3 and Table 4 for their geographic boundaries). Figure 6 shows time series plots of the daily mean HCHO slant column densities and the relative root mean square (RMS) of radiance fitting residual, and Figure 7 shows time series plots of the daily mean HCHO vertical column densities and air mass factors for Collection 3 (left panels) and Collection 4 (right panels). In Collection 4 time series, pixels with main data quality flags labeled as bad, solar zenith angles above 70°, and cloud fractions larger than 0.4 are excluded. In the Collection 3 time series, the same filtering criteria are applied along with an additional exclusion criterion for absolute slant column amounts larger than 3×10^{17} molecules cm^{-2}. This exclusion is necessary due to the large number of outliers present in Collection 3 retrievals after 2012. These outliers are bad retrievals, and their inclusion prevented meaningful interpretation of daily average results (Zhu et al., 2016).

The time series comparison in Figures 6 and 7 clearly shows that Collection 4 SCDs and VCDs are more stable over time than those of Collection 3. Both Collection 3 and Collection 4 time series plots display similar seasonal

10.1029/2024EA003792

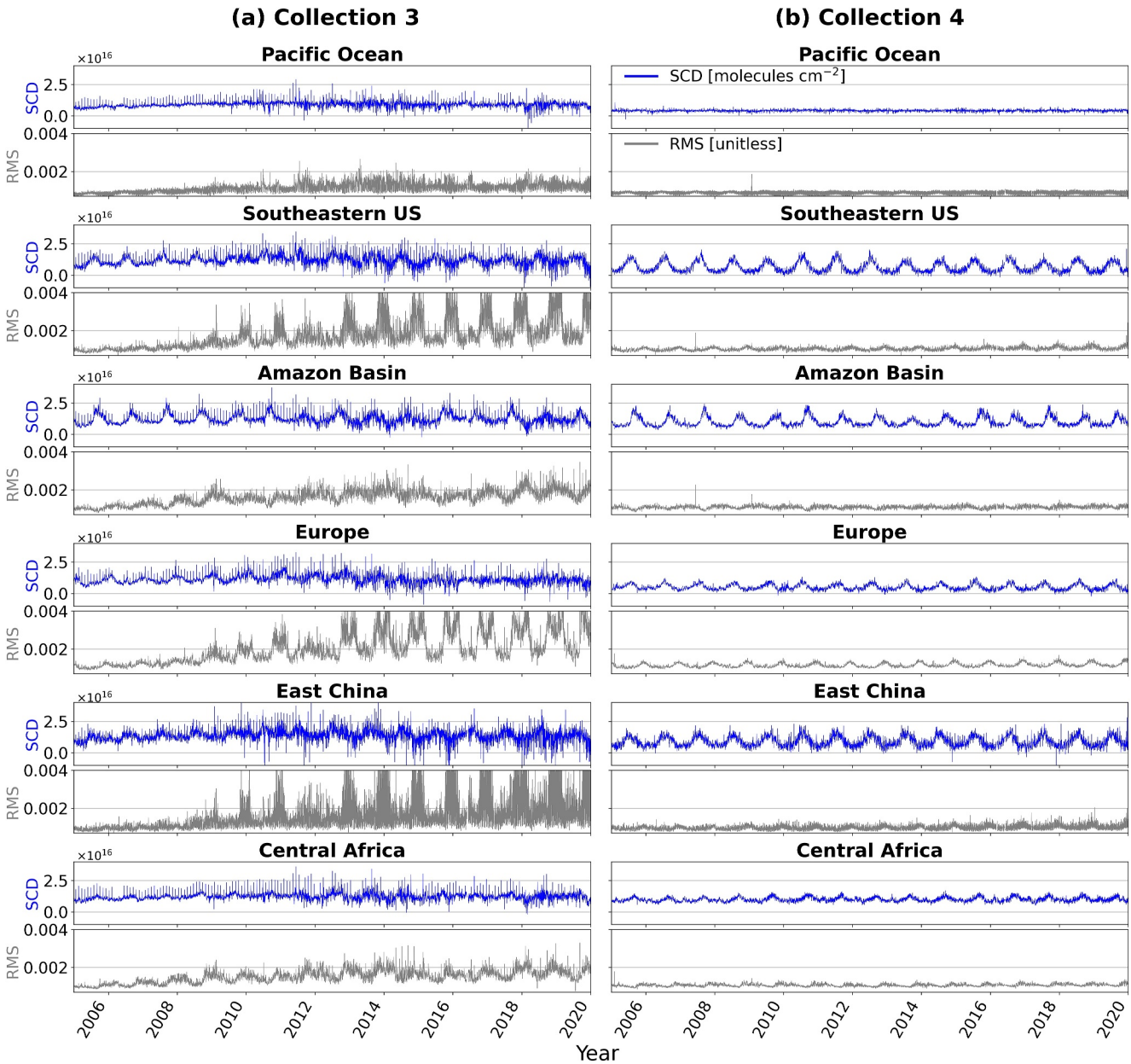

Figure 6. Time series of daily mean HCHO slant column densities and fitting residuals (RMS) for (a) Collection 3, and (b) Collection 4 OMI HCHO for 6 out of 12 regions illustrated in Figure 3.

variability. However, a notable difference arises in the noisy pattern observed in the daily mean SCDs and VCDs of Collection 3, which is no longer present in Collection 4. This is despite the exclusion of pixels with outlier SCDs in the Collection 3 time series. In addition, due to the improvements in Collection 4, the background HCHO column over the Pacific remains constant in time, whereas it fluctuates daily in the Collection 3 time series. Furthermore, the degradation observed in the RMS fitting residuals of Collection 3 is no longer present in Collection 4. Figure 7 shows the daily mean time series of air mass factors for both Collection 3 and Collection 4. The AMFs changes between Collections 3 and 4 reflect the multiple updates in the AMF calculation inputs shown in Table 2. The enhanced stability and reduced noise in Collection 4 daily mean SCD and VCD time series highlight the improved data quality and reliability of this updated data set compared to Collection 3.

Figure 8 shows HCHO slant column densities, air mass factors, and vertical column densities for both Collection 3 (a), Collection 4 (b) and their relative differences (c), all averaged over June 2006 and regridded to 0.1 ° × 0.1°

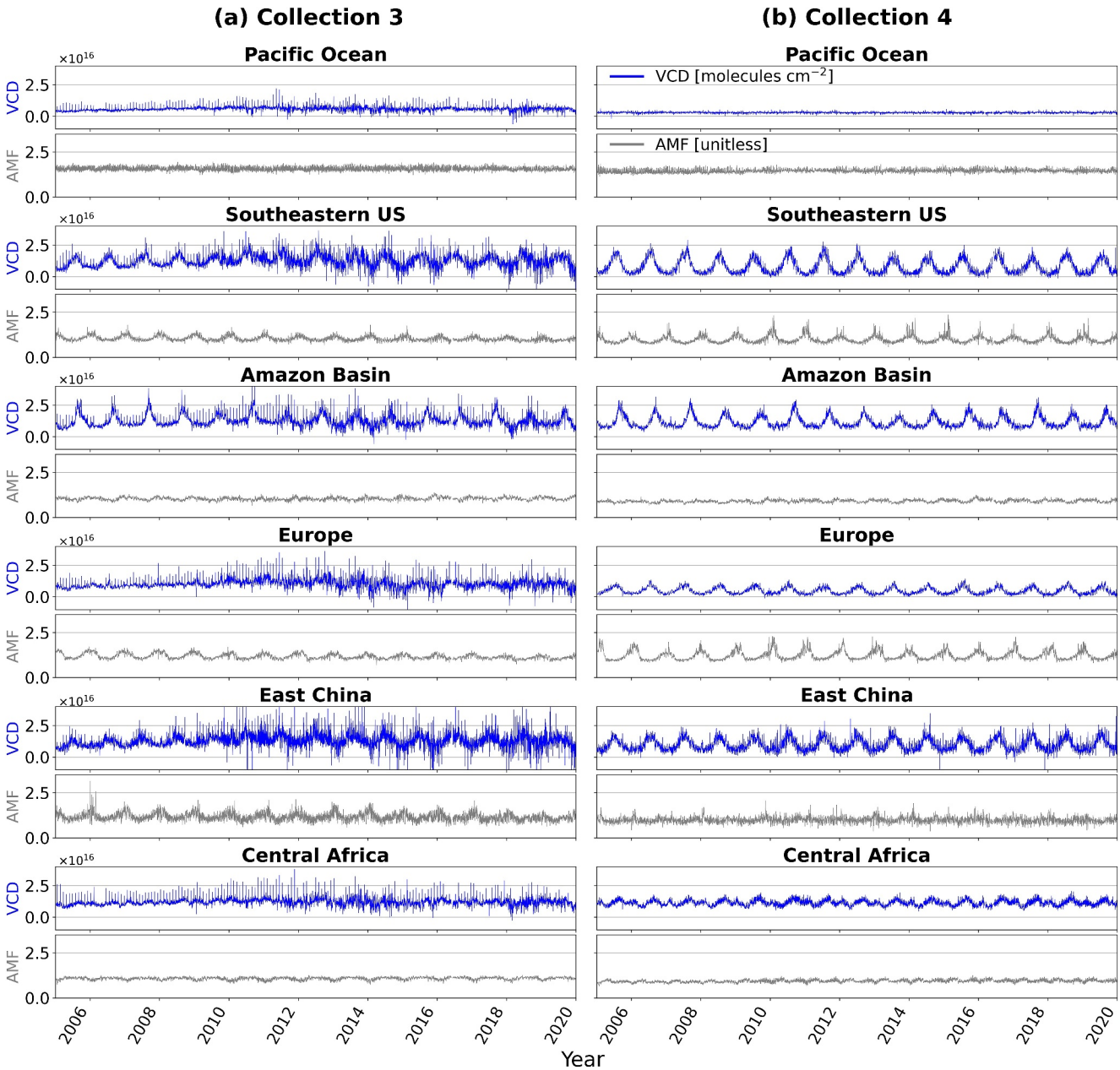

Figure 7. Time series of daily mean HCHO vertical column densities and air mass factors for (a) Collection 3, and (b) Collection 4 OMI HCHO for 6 out of 12 regions illustrated in Figure 3.

over the contiguous United States (CONUS). The data used to generate the plot excludes pixels with solar zenith angles above 70°, cloud fractions greater than 0.4, and main data quality flags labeled as bad. In Figure 8, Collection 3 and Collection 4 data products exhibit a consistent spatial distribution pattern for slant column densities and vertical column densities over CONUS. In this figure, in comparison to Collection 3, the Collection 4 slant column densities over CONUS are smaller by 29% and the Collection 4 vertical column densities are larger by 39%. As discussed in Section 3.1.2, the air mass factor calculations in Collection 4 benefit from ancillary inputs with finer spatial resolution compared to Collection 3 as evident in Figure 8b. The spatial variations observed in air mass factor calculations within Collection 4 compared to Collection 3 can be attributed to two key factors. First, the use of updated ancillary inputs in the radiative transfer model, summarized in Table 2, plays a significant role. Second, the adoption of an updated calculation approach based on scene-by-scene radiative transfer calculation, instead of using look-up tables as in Collection 3, further contributes to these variations. Using the

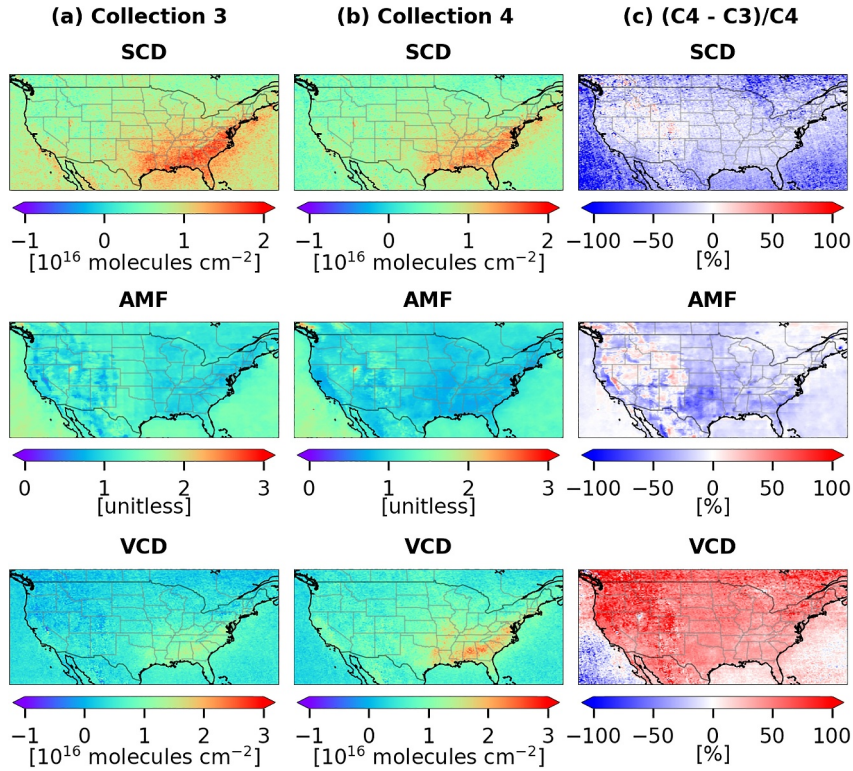

Figure 8. June 2006 mean HCHO slant column densities, air mass factors, and vertical column densities for (a) Collection 3, (b) Collection 4, and (c) relative difference between the two data sets. Mean values are calculated for pixels with solar zenith angle $< 70°$ and cloud fractions < 0.4, regridded to $0.1° \times 0.1°$ resolution.

data presented in Figure 8, we calculated the correlations between OMI Collection 3 and Collection 4 SCDs, AMFs, and VCDs over CONUS as follows: 0.92, 0.84, and 0.85 which show how the spatial distributions of Collection 4 are not drastically different from Collection 3.

4. Validation of the OMI HCHO Product

The OMI Collection 3 HCHO product by González Abad et al. (2015) has previously been validated using observations from aircraft campaigns (Zhu et al., 2016, 2020). Zhu et al. (2016) found a −37% bias in the OMI Collection 3 HCHO product based on validation with aircraft observations and recommended a uniform correction factor of 1.59 to address this issue. Studies that used OMI Collection 3 HCHO product without accounting for this bias may have drawn misleading conclusions. In addition, Zhu et al. (2020) found that OMI Collection 3 HCHO shows an underestimation bias (ranging from 21.7% to 44.5%) under high-HCHO conditions and an overestimation bias (ranging from 66.1% to 112.1%) under low-HCHO conditions. In this study, we validate OMI Collection 4 HCHO product using ground-based FTIR HCHO observations at 27 stations following the validation approach for the OMPS Suomi NPP and OMPS NOAA-20 products by Kwon et al. (2023). We include three more FTIR stations in the Arctic region, Eureka, Ny-Ålesund, and Thule, which were excluded in the OMPS HCHO validation due to the unreliability of the OMPS products over ice and snow.

Figure 9a displays the global distribution of the FTIR stations that are used in this study, providing geolocation and altitude details for each station. These stations represent a diverse set of observation circumstances, ranging from clean remote locations to areas with elevated levels of HCHO pollution. Figure 9b provides the operational timeframe of FTIR stations prior to 2020 which is used in this study. Most of these stations are participants of the NDACC network. The NDACC FTIR HCHO retrievals are harmonized by using consistent spectroscopic parameters and fitting window (Vigouroux et al., 2018). Random and systematic uncertainties in FTIR HCHO measurements are reported as 2.9×10^{14} molecules cm^{-2} (7.6%) and 13.5%, respectively (Vigouroux et al., 2018).

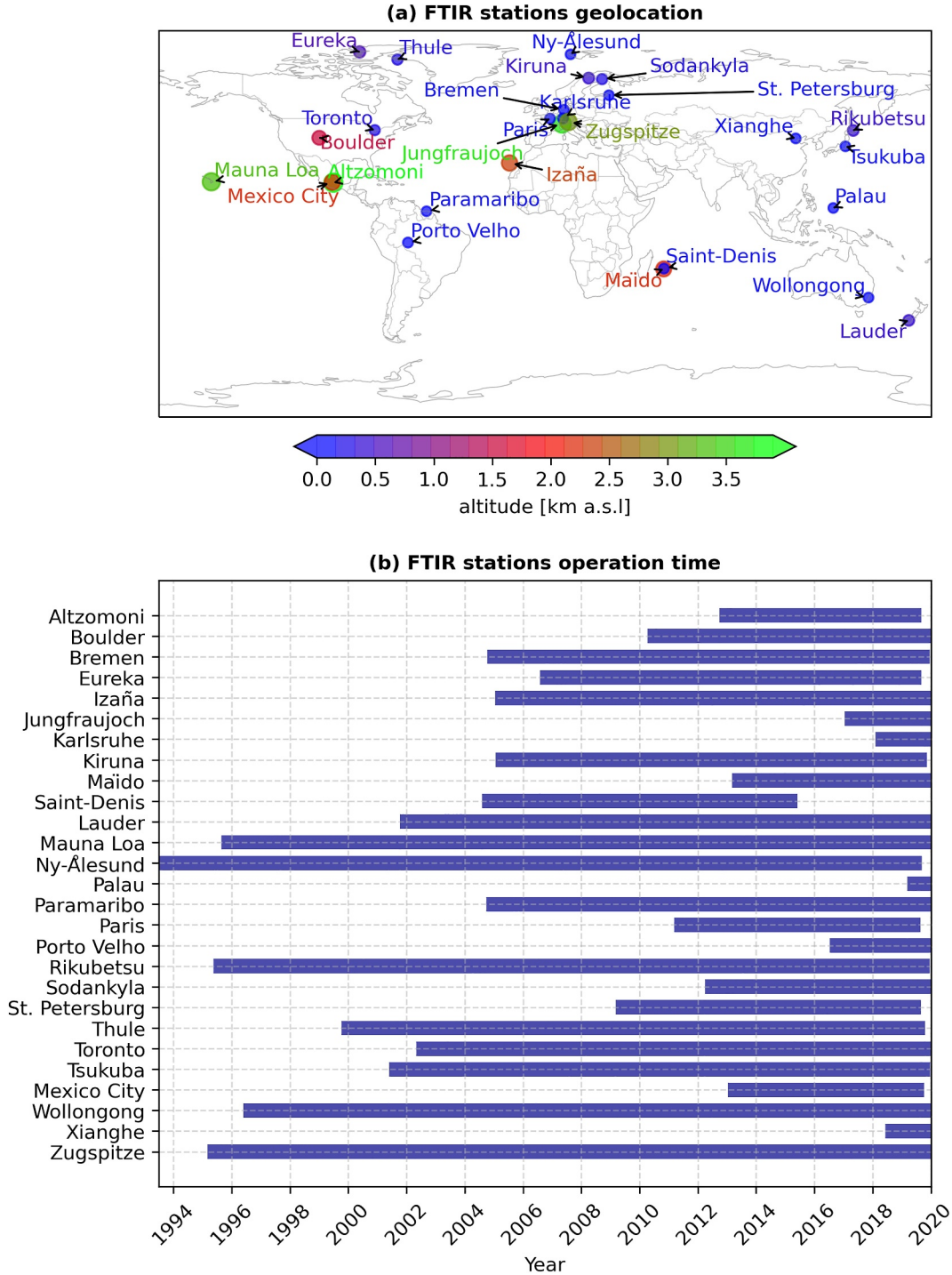

Figure 9. (a) Locations of FTIR stations, and (b) FTIR station operation time periods prior to 2020 which are used in this study.

A detailed description of this validation approach has been explained elsewhere (Kwon et al., 2023; Vigouroux et al., 2020). Here, we provide a brief description of the approach. In order to conduct a validation study, first, we filter OMI HCHO data based on the main data quality flag, solar zenith angle, and cloud fraction. Our validation algorithm only uses the OMI pixels with a main data quality flag of zero, solar zenith angle smaller than 70° and cloud fraction below 0.4. Then, we create pairs of OMI and FTIR observations and generate comparable sets of

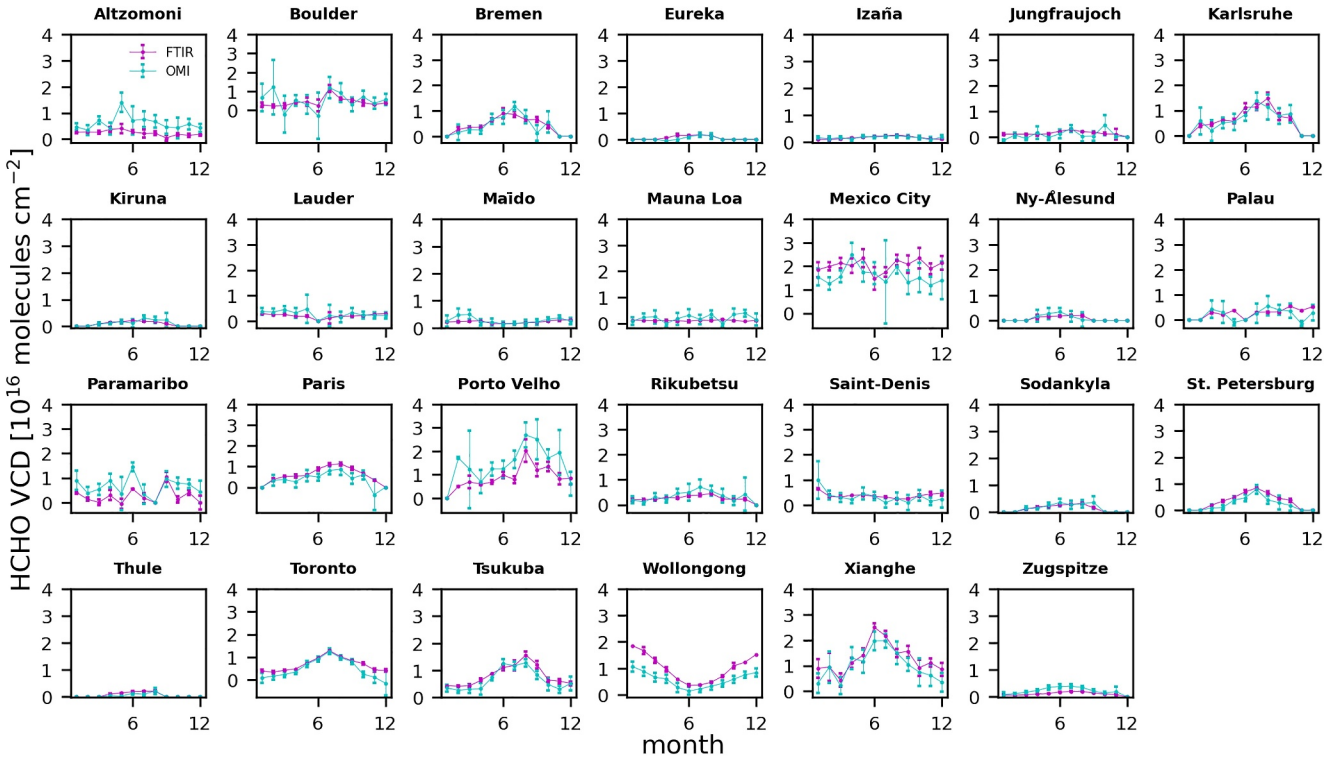

Figure 10. Seasonal variation of HCHO VCDs observed by FTIR (purple) and OMI (blue) at all FTIR stations averaged over 2004–2019. Error bars represent the standard uncertainty described by Kwon et al. (2023). The plot is produced based on climatological monthly mean pairs.

data averaged on a daily, monthly, and climatological monthly basis. We obtain the monthly averaged HCHO by averaging all HCHO observations within each month. Additionally, we explore a scenario where we average all HCHO observations for each calendar month across all years of measurements, aiming to analyze the seasonal patterns. This case is referred to as climatological monthly averaged HCHO.

For each FTIR station, we define a grid with dimensions of 0.2° centered around the station's geographical coordinates. OMI pixels falling within this boundary are averaged. Our choice of 0.2° grid size is driven by both OMI spatial resolution and the necessity of an adequate number of OMI pixels to construct reliable monthly and daily averaged collocated pairs. The OMI HCHO validation incorporates FTIR data measured within a time window of ±3 hr from the OMI observation. The validation utilizes monthly (daily) averaged collocated pairs only when a minimum of 10 (5) OMI pixels contribute to the average. This requirement aims to compensate for OMI HCHO random uncertainty and reduce satellite observation noise. In addition, when comparing averaged collocated OMI and FTIR pairs, we check whether at least the minimum number of 4 pairs exists for each station in order to present meaningful statistics. If this requirement is not satisfied, we exclude those stations from statistical calculations.

Furthermore, we correct for the effects of different a priori profiles and averaging kernels used in FTIR and OMI retrievals following the method explained by Vigouroux et al. (2020) and Kwon et al. (2023). We first interpolate the OMI a priori profile on the FTIR retrieval grid and correct the FTIR retrieved profile using the re-gridded OMI a priori profile (Langerock et al., 2015; Rodgers & Connor, 2003). Then, we construct comparable OMI and FTIR HCHO collocated pairs by interpolating the recalculated FTIR profiles on OMI retrieval grids, smoothing them by applying OMI averaging kernels and scaling them by using a scale factor reflecting the partial HCHO column between the altitude of the satellite pixel and the altitude of the FTIR station (Kwon et al., 2023; Vigouroux et al., 2020).

Figure 10 provides a comparison of the seasonal variability of OMI and FTIR HCHO VCDs for all FTIR stations. We use the climatological monthly averaged collocated pairs to produce these plots. Blue lines show the annual pattern in OMI HCHO VCDs, while purple lines illustrate the annual pattern in FTIR HCHO observations. The

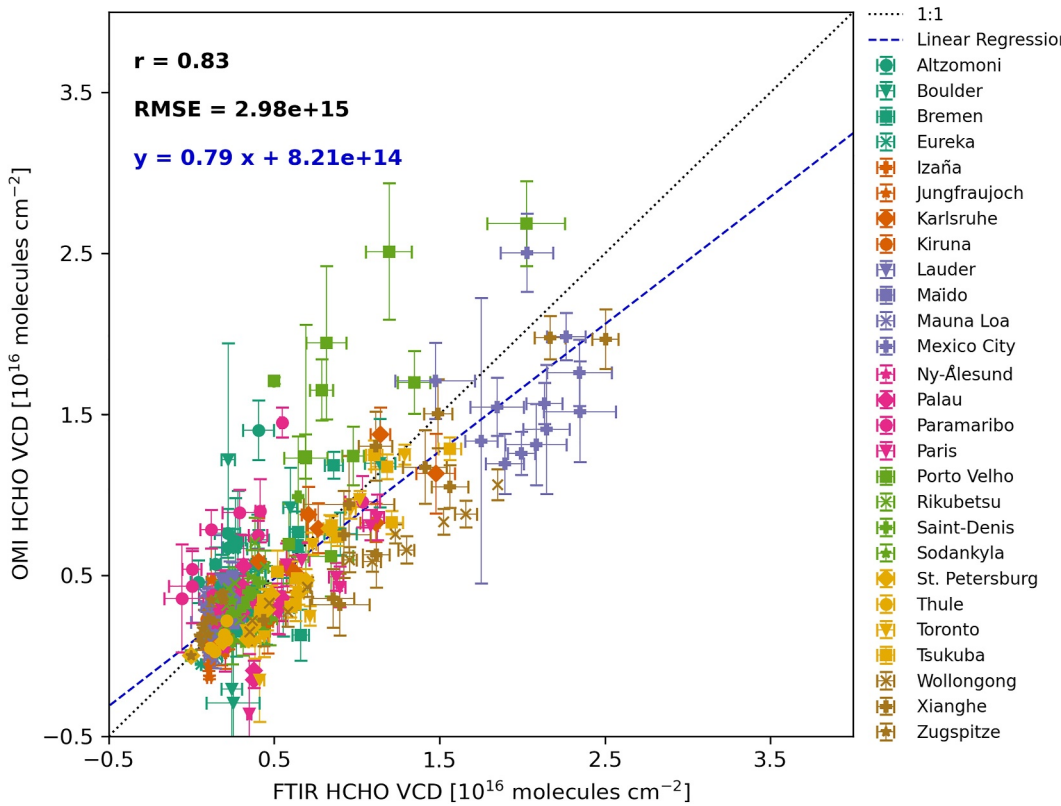

Figure 11. Scatter plot between climatological monthly averaged OMI and FTIR HCHO VCDs. The Pearson correlation coefficient (r), RMSE, regression slope, and intercept are shown on the plot. The blue dashed line shows the regression line.

error bars show their associated standard uncertainties. The standard uncertainty is calculated following the approach discussed by Kwon et al. (2023). The seasonal variability in OMI observations aligns well with FTIR observations. Specifically, at most FTIR stations, we observe a decrease in HCHO VCDs during the winter months followed by an increase during the summer season. The inverse seasonal patterns in the Northern and Southern Hemispheres should be noted. The rise in HCHO VCDs during Summer can be attributed to a larger quantity of HCHO originating from biogenic sources and photochemical reactions occurring in the presence of sunlight. However, this seasonal pattern is not evident in HCHO VCDs from some tropical FTIR stations situated at Altzomoni, Mauna Loa, Mexico City, Palau, and Paramaribo. This result is consistent with those reported by Kwon et al. (2023). It should also be noted that the data at Palau is available over a shorter period, and so is less representative regarding the seasonal pattern than the other stations.

The correlation between climatological monthly averaged pairs of OMI HCHO VCDs and FTIR HCHO VCDs are provided in Figure 11. Each scatter point represents one climatological monthly averaged collocated pair, and horizontal and vertical error bars show the standard uncertainty of FTIR and OMI observations, respectively. The OMI and FTIR HCHO VCDs show a good correlation with each other. Including all climatological monthly averaged pairs from all stations, we obtain a Pearson correlation coefficient (r) and root-mean-square error (RMSE) of 0.83 and 2.98×10^{15} molecules cm^{-2}, respectively. Using linear regression, we determine a regression slope of 0.79 and intercept of 8.21×10^{14} molecules cm^{-2}. The regression line is shown in Figure 11 with a blue dashed line.

In Figure 12a, we compare monthly averaged collocated pairs of OMI and FTIR observations ($r = 0.75$, RMSE = 3.59×10^{15} molecules cm^{-2}, slope = 0.66, and intercept = 1.06×10^{15} molecules cm^{-2}). For context, the correlation analysis between monthly averaged TROPOMI and FTIR HCHO VCDs yielded $r = 0.91$, slope = 0.63, and intercept = 1.17×10^{15} molecules cm^{-2} as reported by Vigouroux et al. (2020). For OMPS Suomi NPP (OMPS NOAA-20), the corresponding statistical parameters were $r = 0.83$ (0.88), slope = 0.82 (0.92), and intercept = 5.71×10^{14} molecules cm^{-2} (6.76×10^{14} molecules cm^{-2}) as determined by Kwon

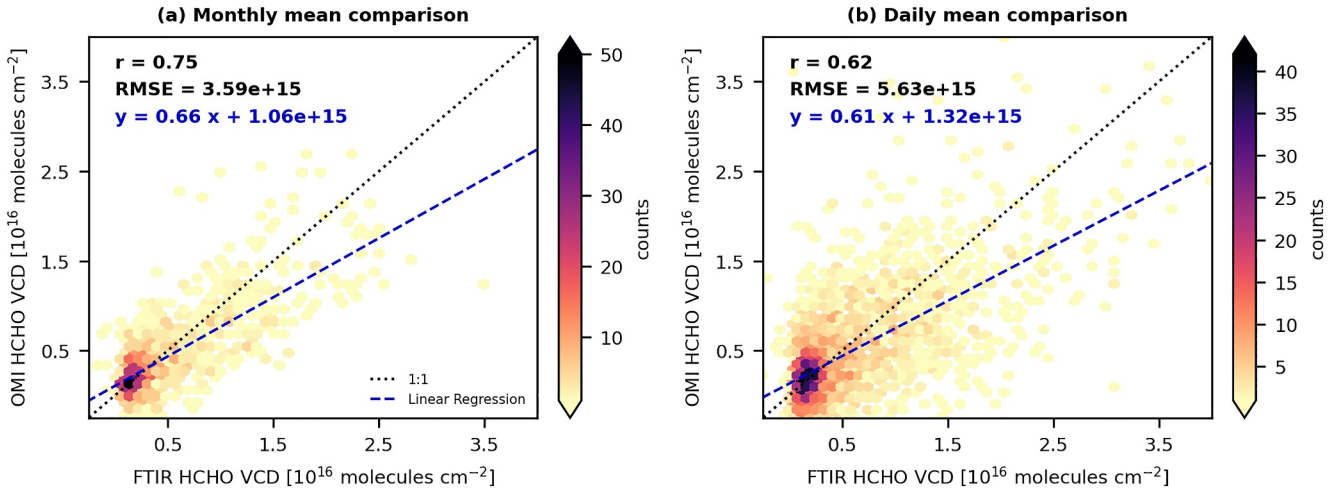

Figure 12. Correlation between (a) monthly averaged, and (b) daily averaged OMI and FTIR HCHO VCDs. The Pearson correlation coefficient (r), RMSE, regression slope, and intercept are shown as text on the plot. The blue dashed line shows the regression line.

et al. (2023). In Figure 12b, we show daily averaged collocated pairs of HCHO columns from OMI and FTIR ($r = 0.62$, RMSE = 5.63×10^{15} molecules cm^{-2}, slope = 0.61, and intercept = 1.32×10^{15} molecules cm^{-2}). Degradation in monthly and daily averaged statistics is expected due to the satellite observation noise compared to the statistics obtained from climatological monthly averaged pairs.

For each station, we determine the overall bias between OMI and FTIR HCHO VCDs using monthly averaged collocated pairs. This bias together with its associated uncertainty is plotted in Figure 13 for all FTIR stations. We calculate the biases and uncertainties following the approach presented by Vigouroux et al. (2020) and Kwon et al. (2023). We excluded two FTIR stations (Jungfraujoch and Palau) from this figure as only one or two pairs remained available for these stations after conducting quality control. As shown in Figure 13, at most FTIR stations, the bias between OMI and FTIR observations falls within ±50%. However, Altzomoni, Boulder, Mauna Loa, Porto Velho, and Zugspitze show higher biases.

As shown in Figures 10 and 13, OMI consistently registers positive biases, that is, higher levels of HCHO VCDs compared to FTIR observations, at elevated clean stations such as Altzomoni, Mauna Loa, and Zugspitze throughout the year. One reason for this discrepancy could be the influence of surrounding HCHO enhancement at lower altitudes within the 0.2° distance of the FTIR locations present in the large footprint of OMI pixels. Conversely, OMI exhibits negative biases, that is, lower VCDs than FTIR, at stations situated in urban areas like Mexico City, Paris, St. Petersburg, and Toronto. This is again likely due to the OMI's larger ground footprint, which includes less-polluted areas located farther from the urban pollution sources. With the higher resolution of future and existing geostationary satellites, we anticipate that the agreement between satellite observations and ground measurements improves.

When comparing OMI and FTIR HCHO columns, the overall mean bias including all stations shown in Figure 13 is 7%. The relatively cleaner stations (those with VCDs smaller than 4×10^{15} molecules cm^{-2}) show a positive bias of 20%, while stations with averaged HCHO VCDs greater than 4×10^{15} molecules cm^{-2}, demonstrate an averaged negative bias of −8%. This pattern of high bias over cleaner stations and low bias over polluted stations is consistent with results from FTIR comparisons with OMPS (Kwon et al., 2023) and TROPOMI (Vigouroux et al., 2020). The comparisons at Porto Velho deviate from the general behavior of biases described above, showing a positive bias of 53% despite FTIR HCHO VCDs exceeding 4×10^{15} molecules cm^{-2}. This result aligns with the OMPS HCHO validation results by Kwon et al. (2023).

5. Intercomparisons With OMPS Instruments and TROPOMI

We intercompare OMI HCHO with OMPS Suomi NPP (Nowlan et al., 2023), OMPS NOAA-20 (Nowlan et al., 2023), and TROPOMI (De Smedt et al., 2018, 2021) HCHO data products. OMI, OMPS Suomi NPP, and OMPS NOAA-20 HCHO data products are produced by the SAO's retrieval algorithm following the same

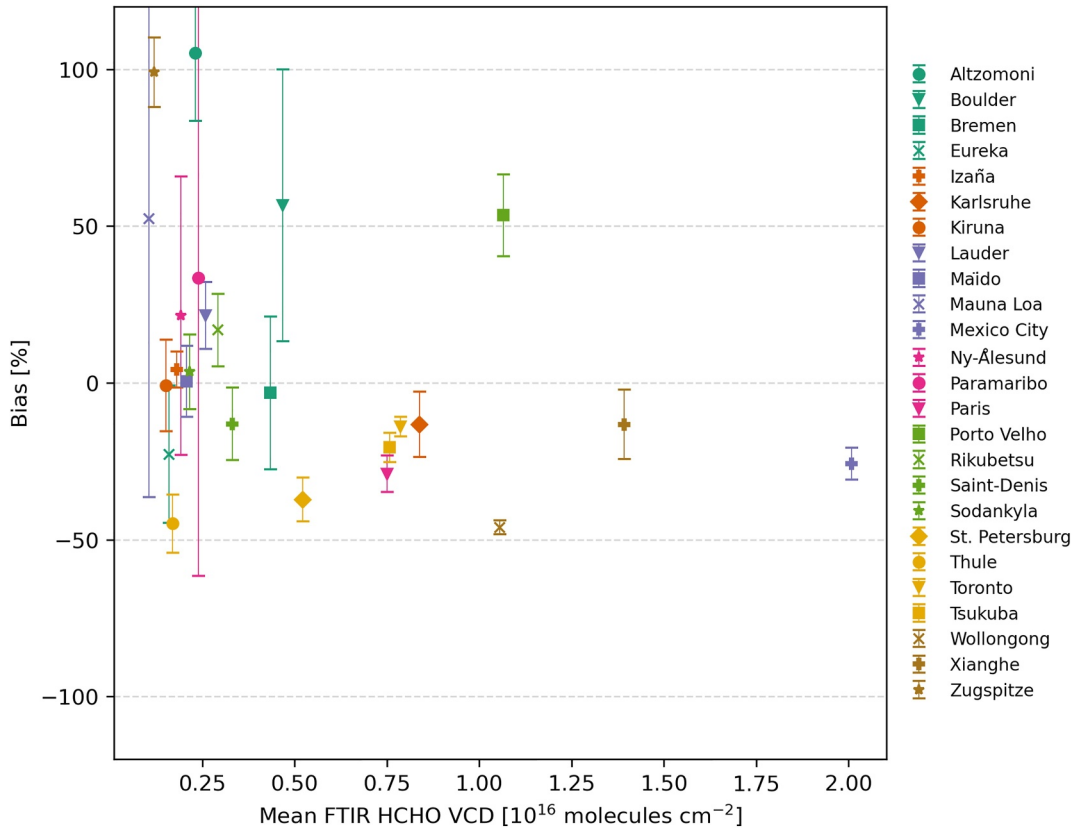

Figure 13. Overall bias between OMI and FTIR observations based on all monthly averaged collocated pairs. Two FTIR stations (Jungfraujoch and Palau) are excluded from this figure due to their small number of collocated pairs after quality control.

pipeline developed within the framework of the MEaSUREs program and using the same fitting inputs. The TROPOMI HCHO data product is retrieved by the BIRA-IASB.

The time series of monthly mean vertical column densities retrieved from OMI (Collection 4), OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI are shown in Figure 14 for 12 regions. The regions of interest are selected following the satellite intercomparison section in the work by Nowlan et al. (2023). These regions are displayed in Figure 3 by blue boxes with geographic boundaries specified in Table 4. In order to calculate the monthly mean VCDs, only pixels with solar zenith angles smaller than 70° and cloud fractions below 0.4 are used. Additionally, pixel quality flags of OMI (main data quality flag = 0), OMPS (main data quality flag = 0), and TROPOMI (quality assurance value ≥0.5) are used as another criterion to ensure data reliability.

In Figure 14, the monthly mean HCHO VCD time series from OMI, OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI demonstrate similar trends and effectively capture the seasonal variations in HCHO column amounts across all 12 geographical regions of interest. Figure 15 provides correlation plots between monthly mean HCHO VCDs from OMI and those from OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI. The monthly means derived from OMI Collection 4 data exhibit high correlations with those from OMPS Suomi NPP and OMPS NOAA-20, with correlation coefficients of 0.98 and 0.97, respectively. The linear regression analysis yields slopes of 0.99 and 0.95, with intercepts of 3.1×10^{14} molecules cm^{-2} and 4.9×10^{14} molecules cm^{-2} between OMI and OMPS Suomi NPP, and OMI and OMPS NOAA-20, respectively. Moreover, the monthly mean HCHO VCDs from OMI and TROPOMI also exhibit a high correlation, with a coefficient of 0.90 while the slope and intercept between them are calculated as 0.67 and 2.3×10^{15} molecules cm^{-2}, respectively.

As shown in Figures 14 and 15, the monthly means from OMI agree very well with OMPS instruments and TROPOMI with high correlation coefficients and small differences compared with the uncertainties in the retrievals. In certain months or regions, there are some disparities or differences but overall, HCHO retrieved from

10.1029/2024EA003792

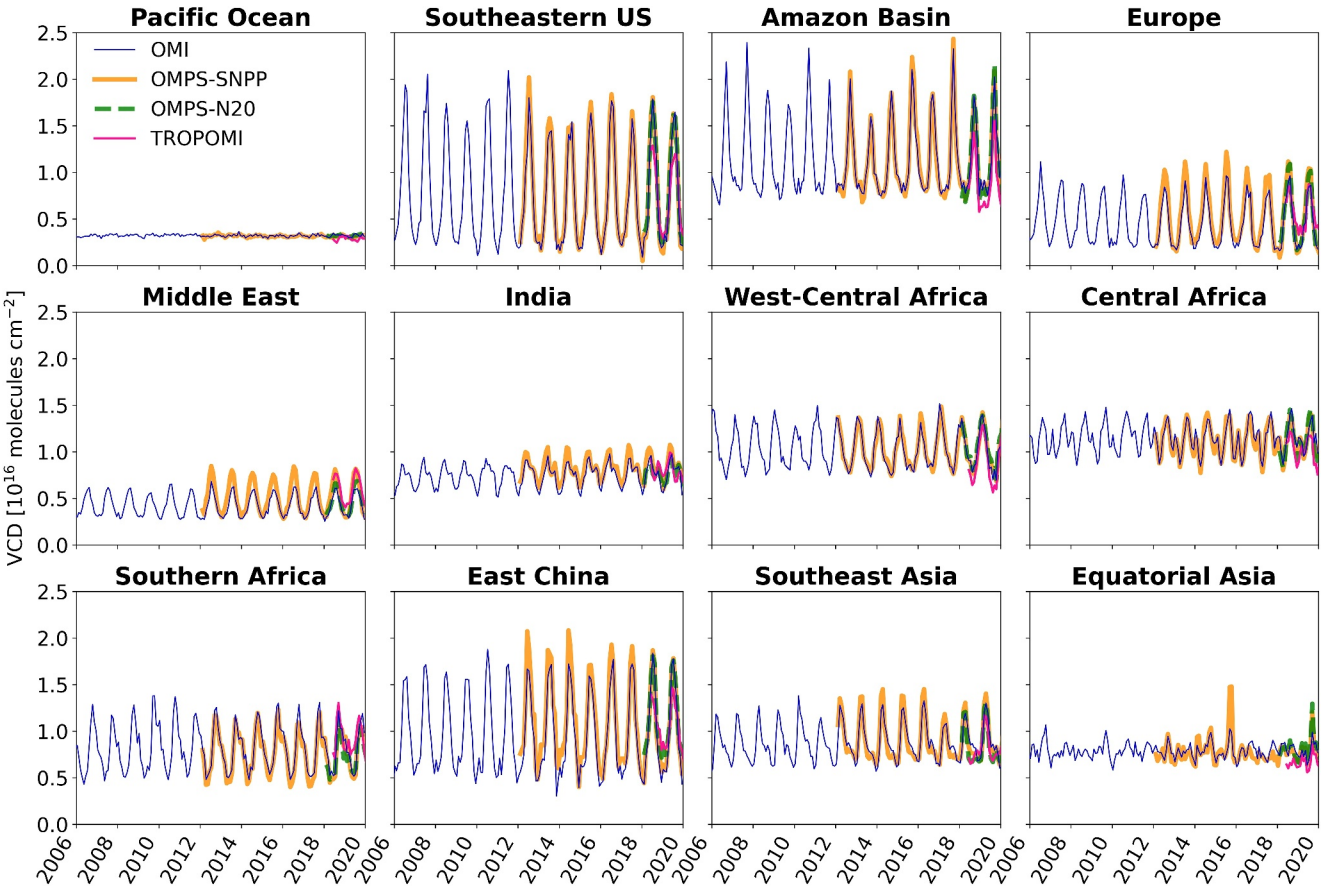

Figure 14. Time series of monthly mean HCHO columns from OMI, OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI for the geographical regions shown in Figure 3.

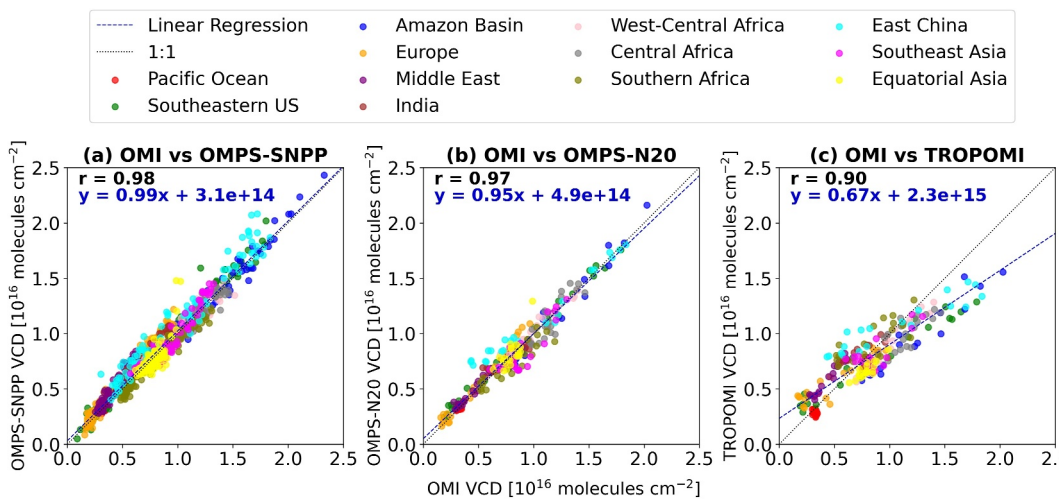

Figure 15. Correlation plots of monthly mean HCHO columns from OMI compared with (a) OMPS Suomi NPP, (b) OMPS NOAA-20, and (c) TROPOMI for the geographical regions shown in Figure 3.

Table 7
Statistics Including Median of Relative Difference (Med. Rel. Diff.), Slope, Intercept, and Pearson Correlation Coefficient (r) of Monthly Mean OMI HCHO Columns Shown in Figure 14 Compared With OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI for All 12 Geographical Regions Shown in Figure 3

	OMI versus OMPS-SNPP					OMI versus OMPS-N20				OMI versus TROPOMI			
	OMI mean VCD (molec cm^{-2})	Med. rel. diff. (%)	Slope	Intercept (molec cm^{-2})	r	Med. rel. diff. (%)	Slope	Intercept (molec cm^{-2})	r	Med. rel. diff. (%)	Slope	Intercept (molec cm^{-2})	r
Pacific Ocean	3.2×10^{15}	0.7	0.47	16.7×10^{14}	0.52	−1.3	0.39	19.9×10^{14}	0.48	12.1	−0.48	44.1×10^{14}	−0.29
Southeastern US	7.6×10^{15}	−4.9	1.04	1.3×10^{14}	0.99	−5.2	0.95	8.3×10^{14}	1	3.5	0.61	24.6×10^{14}	0.98
Amazon Basin	11.4×10^{15}	1.0	1.04	-4.8×10^{14}	0.99	5.0	1.08	-12.0×10^{14}	0.99	29.9	0.76	2.6×10^{14}	0.96
Europe	4.6×10^{15}	−17.6	1.22	-1.6×10^{14}	0.99	−8.2	1.14	-1.4×10^{14}	0.99	−9.1	0.64	26.0×10^{14}	0.93
Middle East	4.2×10^{15}	−16.5	1.45	-10.7×10^{14}	0.96	−8.0	1.08	0.7×10^{14}	0.98	−25.1	1.18	7.2×10^{14}	0.97
India	7.4×10^{15}	−7.6	0.99	7.1×10^{14}	0.93	1.1	0.63	28.9×10^{14}	0.91	−1.3	0.40	49.3×10^{14}	0.53
West-Central Africa	10.3×10^{15}	−0.8	0.95	5.4×10^{14}	0.96	−2.2	0.90	13.1×10^{14}	0.98	13.3	1.06	-16.7×10^{14}	0.98
Central Africa	11.4×10^{15}	1.2	1.01	-2.1×10^{14}	0.95	3.9	1.12	-16.7×10^{14}	0.89	15.4	0.67	23.5×10^{14}	0.71
Southern Africa	8.2×10^{15}	6.6	0.96	-1.6×10^{14}	0.98	8.2	0.83	7.6×10^{14}	0.93	−10.0	0.53	48.5×10^{14}	0.72
East China	9.5×10^{15}	−10.3	1.01	10.8×10^{14}	0.98	−14.6	0.79	32.4×10^{14}	0.98	−10.4	0.44	61.6×10^{14}	0.94
Southeast Asia	8.7×10^{15}	0.5	1.12	-10.6×10^{14}	0.95	4.5	0.98	-4.3×10^{14}	0.88	11.6	0.65	24.4×10^{14}	0.76
Equatorial Asia	7.9×10^{15}	2.9	1.27	-22.1×10^{14}	0.75	0.1	1.20	-14.2×10^{14}	0.74	18.6	0.79	4.6×10^{14}	0.87
All regions	7.9×10^{15}	−2.0	0.99	3.1×10^{14}	0.98	−1.6	0.95	4.9×10^{14}	0.97	7.6	0.67	23.3×10^{14}	0.90

all four instruments displays very high similarities, which highlights the suitability of the OMI Collection 4 HCHO product to perform long-term analysis in combination with other sensors, particularly OMPS Suomi NPP and NOAA-20. The difference between OMI/OMPS and TROPOMI monthly means mainly originates from the differences in the air mass factors as discussed by Nowlan et al. (2023). In addition, multiple factors can explain the remaining bias between the four retrievals including instrument characteristics, overpass time, and cloud property retrieval.

The comparison provided in Table 7 evaluates the monthly mean HCHO VCDs between OMI and OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI, across all geographical regions presented in Figure 14. Metrics such as the median of relative differences (med. rel. diff.), slope, intercept, and Pearson correlation coefficient (r) are provided for each region. The median relative differences between OMI and the other instruments vary across regions: −17.6%–6.6% for OMPS Suomi NPP, −14.6%–8.2% for OMPS NOAA-20%, and -25.1%–29.9% for TROPOMI. However, when aggregating all regions, the monthly mean HCHO values from OMI are overall 2% lower than those from OMPS Suomi NPP, 1.6% lower than those from OMPS NOAA-20%, and 7.6% higher than those from TROPOMI.

Through a linear regression analysis and when excluding the Pacific Ocean, the slope (intercept) between OMI and the other satellite instruments at different geographical regions varies between 0.95 to 1.45 (-22.1×10^{14} to 10.8×10^{14} molecules cm^{-2}) for OMPS Suomi NPP, 0.63 to 1.20 (-16.7×10^{14} to 32.4×10^{14} molecules cm^{-2}) for OMPS NOAA-20, and 0.40 to 1.18 (-16.7×10^{14} to 61.6×10^{14} molecules cm^{-2}) for TROPOMI. Excluding data from both the Pacific Ocean and Equatorial Asia, the Pearson correlation coefficients range from 0.93 to 0.99 for OMI versus OMPS Suomi NPP, 0.88 to 1 for OMI versus OMPS NOAA-20, and 0.53 to 0.98 for OMI versus TROPOMI. The presence of very low HCHO VCDs over the Pacific Ocean together with the absence of large seasonal variations influences the linear regression and correlation results. When comparing OMI with OMPS Suomi NPP, OMPS NOAA-20 and TROPOMI over the Pacific, the slopes are 0.47, 0.39, and −0.48, the

intercepts are 16.7 × 10^{14} molecules cm^{-2}, 19.9 × 10^{14} molecules cm^{-2}, and 44.1 × 10^{14} molecules cm^{-2}, and the correlation coefficients are 0.52, 0.48, and −0.29, respectively.

In Equatorial Asia, OMPS Suomi NPP and OMPS NOAA-20 retrievals capture higher HCHO concentrations linked to large emissions from biomass burning occurring in September 2015 and September 2019, as evidenced by two peaks in Figure 14. However, OMI and TROPOMI exhibit lower HCHO VCDs during those months, likely due to cloud effects on retrievals in case of fires. This discrepancy impacts the correlation coefficient between OMI and OMPS instruments over this region, resulting in correlation coefficients of 0.75 (OMPS Suomi NPP) and 0.74 (OMPS NOAA-20). The cloud data used in the OMPS, OMI and TROPOMI retrievals show notable differences. Applying a cloud fraction filter of 0.4 across all three instruments leads to filtering out more pixels over areas affected by thick smoke or clouds created from wildfires in the OMI and TROPOMI retrievals. As a consequence, the HCHO columns observed by OMPS appear substantially higher than those observed by OMI and TROPOMI. It is important to acknowledge that the uncertainties in the AMFs under these observation conditions are larger than typical AMF uncertainties due to our limited understanding of the scattering conditions of the atmosphere.

6. Summary

We present the new OMI Collection 4 HCHO product, retrieved using the SAO MEaSUREs algorithm. The retrieval algorithm involves three main steps: calculating differential slant column density, deriving air mass factor, and applying a reference sector correction. In this paper, we discuss the updates in the retrieval algorithm in comparison to the current NASA operational product (OMI Collection 3 HCHO). These updates include the enhancements in Level-1B radiance data, improvements in spectral fitting and radiative transfer modeling and input parameters, and integration of an updated bias correction procedure in the reference sector correction step.

The OMI Collection 4 HCHO product shows great stability, good accuracy, and minimal noise increase over time. In the OMI Collection 4 HCHO product, we estimate median random uncertainties to be on the order of ∼6.5 × 10^{15} in SCDs and ∼5 × 10^{15} molecules cm^{-2} in VCDs, with ∼11 − 14% of the random uncertainty associated with the AMF calculation.

We validate the new HCHO data product using ground-based FTIR measurements at 27 stations. The climatological monthly averaged OMI Collection 4 HCHO VCDs show a good agreement with the FTIR VCDs with a correlation coefficient of 0.83, RMSE of 2.98 × 10^{15} molecules cm^{-2}, regression slope of 0.79, and intercept of 8.21 × 10^{14} molecules cm^{-2}. When compared with ground-based measurements, OMI exhibits an average negative bias of −8% over polluted stations (VCDs >4 × 10^{15} molecules cm^{-2}) and a positive bias of 20% over relatively cleaner stations (VCDs <4 × 10^{15} molecules cm^{-2}). The seasonal variability in OMI observations aligns well with FTIR measurements.

We also compare the OMI Collection 4 HCHO VCDs to HCHO VCDs retrieved from OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI. The monthly averaged OMI HCHO VCDs are highly correlated with OMPS Suomi NPP, OMPS NOAA-20, and TROPOMI in 12 geographic regions with correlation coefficients of 0.98, 0.97, and 0.90, respectively. The monthly mean HCHO VCDs from OMI are lower compared to those from OMPS Suomi NPP and OMPS NOAA-20 by approximately −2% and −1.6%, respectively, but they are higher by 7.6% compared to the corresponding measurements from TROPOMI.

In summary, the OMI Collection 4 HCHO data set demonstrates remarkable stability, making it a reliable resource for establishing a long-term record of HCHO retrieval from space. Its strong agreement with OMPS and TROPOMI HCHO products, specially with OMPS products, makes the OMI Collection 4 product an excellent data set for conducting multidecadal trend and climatological analysis that can be expanded in the future using newer data sets.

Data Availability Statement

The OMI Collection 4 HCHO product is available upon request from the authors. Upon final implementation of the OMI Collection 4 HCHO algorithm in NASA's operational processing system, the product will become available from the NASA GES DISC, substituting OMI Collection 3 HCHO product (Chance, 2007). The OMI Level-1B Radiances (Kleipool, 2021b) is available from the NASA GES DISC. The MODIS BRDF product (MCD43C1) (Schaaf & Wang, 2021) is available through the NASA LP DAAC. MERRA-2 data (GMAO, 2015)

Acknowledgments
This research has been supported by NASA's Making Earth System Data Records for Use in Research Environments (80NSSC18M0091, 80NSSC24M0037), and NASA's Aura Project Core Data Analysis (80NSSC21K0177, 80NSSC24K0120) grants. Computations in this paper were conducted on the Smithsonian High Performance Cluster (SI/HPC), Smithsonian Institution (https://doi.org/10.25572/SIHPC). The National Center for Atmospheric Research is sponsored by the National Science Foundation. The NCAR FTS observation programs at Thule, GR, Boulder, CO and Mauna Loa, HI are supported under contract by the National Aeronautics and Space Administration (NASA). The Thule work is also supported by the NSF Office of Polar Programs (OPP). We wish to thank the Danish Meteorological Institute for support at the Thule site and NOAA for support of the MLO site. The Bremen FTIR measurements have been supported by the BMBF (German Ministry of Research and Education) in the project ROMIC-II subproject TroStra (01LG1904A) and the DFG (Deutsche Forschungsgemeinschaft, German Research Foundation) project 404/27-1. The authors want to thank Patrick Tellei, President of the Palau Community College, for the provision of space for the laboratory containers in the college; German Honorary Consul Thomas Schubert, for overall support; and various people and institutions for operations at the PAO: Katrin Müller (Alfred Wegener Institute, Helmholtz Centre for Polar and Marine Research, AWI), Christoph Ritter (AWI), Jürgen "Egon" Graeser (AWI), Ingo Beninga (Impres GmbH), Wilfried Ruhe (Impres GmbH), Winfried Markert (Uni Bremen), and Sharon Patris (Coral Reef Research Foundation, Palau). The Eureka FTIR measurements were made at the Polar Environment Atmospheric Research Laboratory by the Canadian Network for the Detection of Atmospheric Change, primarily supported by the Natural Sciences and Engineering Research Council of Canada (NSERC), Environment and Climate Change Canada (ECCC) and the Canadian Space Agency. The Toronto FTIR measurements were made at the University of Toronto Atmospheric Observatory, primarily supported by NSERC and ECCC. The FTIR monitoring program at Jungfraujoch was primarily supported by the F.R.S.-FNRS (Brussels, Belgium) and the GAW-CH program of MeteoSwiss (Zürich, Switzerland). E. Mahieu is a senior research associate with F.R.S.-FNRS. Lauder FTIR measurements are core-funded by NIWA through New Zealand's Ministry of Business, Innovation and Employment Strategic Science Investment Fund. The FTIR measurement site in Paris has received funding from Sorbonne Université, the French research center

is available from the NASA GES DISC. The ocean salinity product (Levitus & US NODC, 2013) can be found from the NOAA NCEI. The MODIS Terra chlorophyll product (Werdell et al., 2023) is available at NASA Earth Data. The FTIR data (NDACC, 2023) is available at NASA LARC or under request to the FTIR PIs. Please review the NDACC data use agreement prior to utilizing the data. The OMPS HCHO products used for inter-comparison purposes can be accessed from the NASA GES DISC for OMPS Suomi-NPP (González Abad, 2022b) and OMPS NOAA-20 (González Abad, 2022a) and TROPOMI HCHO product (ESA & DLR, 2019a, 2019b) is available at NASA GES DISC.

References

Anderson, L. G., Lanning, J. A., Barrell, R., Miyagishima, J., Jones, R. H., & Wolfe, P. (1996). Sources and sinks of formaldehyde and acetaldehyde: An analysis of Denver's ambient concentration data. *Atmospheric Environment*, *30*(12), 2113–2123. https://doi.org/10.1016/1352-2310(95)00175-1

Antonov, J., Seidov, D., Boyer, T., Locarnini, R., Mishonov, A., Garcia, H., et al. (2010). In S. Levitus (Ed.), *World ocean atlas 2009 volume 2: Salinity, NOAA atlas NESDIS 69 (Unpublished doctoral dissertation)* (p. 184). US Government Printing Office.

Bak, J., Liu, X., Kim, J.-H., Haffner, D. P., Chance, K., Yang, K., & Sun, K. (2017). Characterization and correction of OMPS nadir mapper measurements for ozone profile retrievals. *Atmospheric Measurement Techniques*, *10*(11), 4373–4388. https://doi.org/10.5194/amt-10-4373-2017

Barkley, M. P., Smedt, I. D., Van Roozendael, M., Kurosu, T. P., Chance, K., Arneth, A., et al. (2013). Top-down isoprene emissions over tropical South America inferred from SCIAMACHY and OMI formaldehyde columns. *Journal of Geophysical Research: Atmospheres*, *118*(12), 6849–6868. https://doi.org/10.1002/jgrd.50552

Beirle, S., Lampel, J., Lerot, C., Sihler, H., & Wagner, T. (2017). Parameterizing the instrumental spectral response function and its changes by a super-Gaussian and its derivatives. *Atmospheric Measurement Techniques*, *10*(2), 581–598. https://doi.org/10.5194/amt-10-581-2017

Bey, I., Jacob, D. J., Yantosca, R. M., Logan, J. A., Field, B. D., Fiore, A. M., et al. (2001). Global modeling of tropospheric chemistry with assimilated meteorology: Model description and evaluation. *Journal of Geophysical Research*, *106*(D19), 23073–23095. https://doi.org/10.1029/2001jd000807

Boersma, K., Eskes, H., Dirksen, R., Van Der A, R., Veefkind, J., Stammes, P., et al. (2011). An improved tropospheric NO_2 column retrieval algorithm for the ozone monitoring instrument. *Atmospheric Measurement Techniques*, *4*(9), 1905–1928. https://doi.org/10.5194/amt-4-1905-2011

Brune, W. H., Tan, D., Faloona, I., Jaeglé, L., Jacob, D. J., Heikes, B., et al. (1999). OH and HO_2 chemistry in the North Atlantic free troposphere. *Geophysical Research Letters*, *26*(20), 3077–3080. https://doi.org/10.1029/1999gl900549

Chance, K. (2007). OMI/Aura Formaldehyde (HCHO) Total Column 1-orbit L2 Swath 13x24 km [Dataset]. *NASA Goddard Earth Sciences Data and Information Services Center*. https://doi.org/10.5067/AURA/OMI/DATA2015

Chance, K., Kurosu, T. P., & Sioris, C. E. (2005). Undersampling correction for array detector-based satellite spectrometers. *Applied Optics*, *44*(7), 1296–1304. https://doi.org/10.1364/ao.44.001296

Chance, K., & Kurucz, R. L. (2010). An improved high-resolution solar reference spectrum for Earth's atmosphere measurements in the ultraviolet, visible, and near infrared. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *111*(9), 1289–1295. https://doi.org/10.1016/j.jqsrt.2010.01.036

Chance, K., & Orphal, J. (2011). Revised ultraviolet absorption cross sections of H_2CO for the HITRAN database. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *112*(9), 1509–1510. https://doi.org/10.1016/j.jqsrt.2011.02.002

Chance, K., Palmer, P. I., Spurr, R. J., Martin, R. V., Kurosu, T. P., & Jacob, D. J. (2000). Satellite observations of formaldehyde over North America from GOME. *Geophysical Research Letters*, *27*(21), 3461–3464. https://doi.org/10.1029/2000gl011857

Chance, K., & Spurr, R. J. (1997). Ring effect studies: Rayleigh scattering, including molecular parameters for rotational Raman scattering, and the Fraunhofer spectrum. *Applied Optics*, *36*(21), 5224–5230. https://doi.org/10.1364/ao.36.005224

Chong, H., González Abad, G., Nowlan, C. R., Chan Miller, C., Saiz-Lopez, A., Fernandez, R. P., et al. (2024). Global retrieval of stratospheric and tropospheric BrO columns from the Ozone Mapping and Profiler Suite Nadir Mapper (OMPS-NM) on board the Suomi-NPP satellite. *Atmospheric Measurement Techniques*, *17*(9), 2873–2916. https://doi.org/10.5194/amt-17-2873-2024

Cox, C., & Munk, W. (1954). Measurement of the roughness of the sea surface from photographs of the Sun's glitter. *Josa*, *44*(11), 838–850. https://doi.org/10.1364/josa.44.000838

De Smedt, I., Müller, J.-F., Stavrakou, T., Van Der A, R., Eskes, H., & Van Roozendael, M. (2008). Twelve years of global observations of formaldehyde in the troposphere using GOME and SCIAMACHY sensors. *Atmospheric Chemistry and Physics*, *8*(16), 4947–4963. https://doi.org/10.5194/acp-8-4947-2008

De Smedt, I., Pinardi, G., Vigouroux, C., Compernolle, S., Bais, A., Benavent, N., et al. (2021). Comparative assessment of TROPOMI and OMI formaldehyde observations and validation against MAX-DOAS network column measurements. *Atmospheric Chemistry and Physics*, *21*(16), 12561–12593. https://doi.org/10.5194/acp-21-12561-2021

De Smedt, I., Stavrakou, T., Hendrick, F., Danckaert, T., Vlemmix, T., Pinardi, G., et al. (2015). Diurnal, seasonal and long-term variations of global formaldehyde columns inferred from combined OMI and GOME-2 observations. *Atmospheric Chemistry and Physics*, *15*(21), 12519–12545. https://doi.org/10.5194/acp-15-12519-2015

De Smedt, I., Theys, N., Yu, H., Danckaert, T., Lerot, C., Compernolle, S., et al. (2018). Algorithm theoretical baseline for formaldehyde retrievals from S5P TROPOMI and from the QA4ECV project. *Atmospheric Measurement Techniques*, *11*(4), 2395–2426. https://doi.org/10.5194/amt-11-2395-2018

De Smedt, I., Van Roozendael, M., Stavrakou, T., Müller, J.-F., Lerot, C., Theys, N., et al. (2012). Improved retrieval of global tropospheric formaldehyde columns from GOME-2/MetOp-A addressing noise reduction and instrumental degradation issues. *Atmospheric Measurement Techniques*, *5*(11), 2933–2949. https://doi.org/10.5194/amt-5-2933-2012

Dirksen, R., Dobber, M., Voors, R., & Levelt, P. (2006). Prelaunch characterization of the ozone monitoring instrument transfer function in the spectral domain. *Applied Optics*, *45*(17), 3972–3981. https://doi.org/10.1364/ao.45.003972

Duncan, B. N., Yoshida, Y., Olson, J. R., Sillman, S., Martin, R. V., Lamsal, L., et al. (2010). Application of OMI observations to a space-based indicator of NO_x and VOC controls on surface ozone formation. *Atmospheric Environment*, *44*(18), 2213–2223. https://doi.org/10.1016/j.atmosenv.2010.03.010

CNRS and the French space agency CNES. The FTIR operations of Rikubetsu and Tsukuba are supported in part by the GOSAT series project. The Rikubetsu NDACC site is funded by the joint research program of the Institute for Space-Earth Environmental Research (ISEE), Nagoya University. The authors acknowledge Saint-Petersburg State University for a research project 123042000071-8 (GZ-MDF-2023-2, PURE ID 93882802). We thank the Porto Velho and Reunion Island FTIR teams for providing ground-based HCHO data.

ESA & DLR. (2019a). Sentinel-5P TROPOMI tropospheric formaldehyde HCHO 1-orbit L2 5.5 km × 3.5km [Dataset]. *Goddard Earth Sciences Data and Information Services Center (GES DISC). (Copernicus Sentinel data processed by ESA, German Aerospace Center (DLR))*. https://doi.org/10.5270/S5P-vg1i7t0

ESA & DLR. (2019b). Sentinel-5P TROPOMI tropospheric formaldehyde HCHO 1-orbit L2 7 km × 3.5 km [Dataset]. *Goddard Earth Sciences Data and Information Services Center (GES DISC). (Copernicus Sentinel data processed by ESA, German Aerospace Center (DLR))*. https://doi.org/10.5270/S5P-vg1i7t0

Fasnacht, Z., Vasilkov, A., Haffner, D., Qin, W., Joiner, J., Krotkov, N., et al. (2019). A geometry-dependent surface Lambertian-equivalent reflectivity product for UV–Vis retrievals–part 2: Evaluation over open ocean. *Atmospheric Measurement Techniques*, *12*(12), 6749–6769. https://doi.org/10.5194/amt-12-6749-2019

Fetterer, F., Knowles, K., Meier, W., Savoie, M., & Windnagel, A. (2017). *Sea ice index, version 3*. National Snow and Ice Data Center.

Finkenzeller, H., & Volkamer, R. (2022). O_2–O_2 CIA in the gas phase: Cross-section of weak bands, and continuum absorption between 297–500 nm. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *279*, 108063. https://doi.org/10.1016/j.jqsrt.2021.108063

Friedl, M., & Sulla-Menashe, D. (2019). MCD12Q1 MODIS/Terra+Aqua land cover type yearly L3 global 500m SIN grid V006 [Dataset]. https://doi.org/10.5067/MODIS/MCD12Q1.006

Gelaro, R., McCarty, W., Suárez, M. J., Todling, R., Molod, A., Takacs, L., et al. (2017). The Modern-Era Retrospective Analysis for Research and Applications, version 2 (MERRA-2). *Journal of Climate*, *30*(14), 5419–5454. https://doi.org/10.1175/JCLI-D-16-0758.1

GEOS-Chem Support Team. (2018). GEOS-Chem online user's guide. Retrieved from https://geoschem.github.io/gcclassic-manpage-archive/man.GC_12/index.html

GMAO. (2015). MERRA-2 tavg1_2d_slv_Nx: 2d,1-hourly,time-averaged,single-level,assimilation,single-level diagnostics V5.12.4 [Dataset]. *Goddard Earth Sciences Data and Information Services Center (GES DISC)*. https://doi.org/10.5067/VJAFPLI1CSIV

González Abad, G. (2022a). OMPS-N20 L2 NM formaldehyde (HCHO) total column swath orbital V1 [Dataset]. *Goddard Earth Sciences Data and Information Services Center (GES DISC)*. https://doi.org/10.5067/CIYXT9A4I2F4

González Abad, G. (2022b). OMPS-NPP L2 NM formaldehyde (HCHO) total column swath orbital V1 [Dataset]. *Goddard Earth Sciences Data and Information Services Center (GES DISC)*. https://doi.org/10.5067/IIM1GHT07QA8

González Abad, G., Liu, X., Chance, K., Wang, H., Kurosu, T., & Suleiman, R. (2015). Updated Smithsonian astrophysical observatory ozone monitoring instrument (SAO OMI) formaldehyde retrieval. *Atmospheric Measurement Techniques*, *8*(1), 19–32. https://doi.org/10.5194/amt-8-19-2015

González Abad, G., Vasilkov, A., Seftor, C., Liu, X., & Chance, K. (2016). Smithsonian astrophysical observatory ozone mapping and profiler suite (SAO OMPS) formaldehyde retrieval. *Atmospheric Measurement Techniques*, *9*(7), 2797–2812. https://doi.org/10.5194/amt-9-2797-2016

Hastings, D. A., Dunbar, P. K., Elphingstone, G. M., Bootz, M., Murakami, H., Maruyama, H., et al. (1999). *The global land one-kilometer base elevation (globe) digital elevation model, version 1.0* (Vol. 325, pp. 80305–83328). National Oceanic and Atmospheric Administration, National Geophysical Data Center.

Howlett, C., González Abad, G., Chan Miller, C., Nowlan, C. R., Ayazpour, Z., & Zhu, L. (2023). The influence of snow cover on ozone monitor instrument formaldehyde observations. *Atmósfera*, *37*. https://doi.org/10.20937/ATM.53134

Hu, C., Lee, Z., & Franz, B. (2012). Chlorophyll aalgorithms for oligotrophic oceans: A novel approach based on three-band reflectance difference. *Journal of Geophysical Research*, *117*(C1), C01011. https://doi.org/10.1029/2011jc007395

Ingmann, P., Veihelmann, B., Langen, J., Lamarre, D., Stark, H., & Courrèges-Lacoste, G. B. (2012). Requirements for the GMES atmosphere service and ESA's implementation concept: Sentinels-4/-5 and-5P. *Remote Sensing of Environment*, *120*, 58–69. https://doi.org/10.1016/j.rse.2012.01.023

Jin, X., Fiore, A. M., Murray, L. T., Valin, L. C., Lamsal, L. N., Duncan, B., et al. (2017). Evaluating a space-based indicator of surface ozone-NO_x-VOC sensitivity over midlatitude source regions and application to decadal trends. *Journal of Geophysical Research: Atmospheres*, *122*(19), 10–439. https://doi.org/10.1002/2017jd026720

Jin, X., & Holloway, T. (2015). Spatial and temporal variability of ozone sensitivity over China observed from the ozone monitoring instrument. *Journal of Geophysical Research: Atmospheres*, *120*(14), 7229–7246. https://doi.org/10.1002/2015JD023250

Jung, Y., González Abad, G., Nowlan, C. R., Chance, K., Liu, X., Torres, O., & Ahn, C. (2019). Explicit aerosol correction of OMI formaldehyde retrievals. *Earth and Space Science*, *6*(11), 2087–2105. https://doi.org/10.1029/2019ea000702

Kaiser, J., Jacob, D. J., Zhu, L., Travis, K. R., Fisher, J. A., González Abad, G., et al. (2018). High-resolution inversion of OMI formaldehyde columns to quantify isoprene emission on ecosystem-relevant scales: Application to the Southeast US. *Atmospheric Chemistry and Physics*, *18*(8), 5483–5497. https://doi.org/10.5194/acp-18-5483-2018

Kim, J., Jeong, U., Ahn, M.-H., Kim, J. H., Park, R. J., Lee, H., et al. (2020). New era of air quality monitoring from space: Geostationary environment monitoring spectrometer (GEMS). *Bulletin of the American Meteorological Society*, *101*(1), E1–E22. https://doi.org/10.1175/bams-d-18-0013.1

Kleipool, Q. (2021a). OMI/Aura level 1B averaged solar irradiances V004 [Dataset]. *Goddard Earth Sciences Data and Information Services Center (GES DISC)*. https://doi.org/10.5067/Aura/OMI/DATA1401

Kleipool, Q. (2021b). OMI/Aura level 1B UV global geolocated Earthshine radiances V004 [Dataset]. *Archived by National Aeronautics and Space Administration, U.S. Government, Goddard Earth Sciences Data and Information Services Center (GES DISC)*. https://doi.org/10.5067/AURA/OMI/DATA1402

Kleipool, Q., Dobber, M., de Haan, J., & Levelt, P. (2008). Earth surface reflectance climatology from 3 years of OMI data. *Journal of Geophysical Research*, *113*(D18), D18308. https://doi.org/10.1029/2008jd010290

Kleipool, Q., Rozemeijer, N., van Hoek, M., Leloux, J., Loots, E., Ludewig, A., et al. (2022). Ozone monitoring instrument (OMI) collection 4: Establishing a 17-year-long series of detrended level-1b data. *Atmospheric Measurement Techniques*, *15*(11), 3527–3553. https://doi.org/10.5194/amt-15-3527-2022

Kurosu, T. P., Chance, K., & Sioris, C. E. (2004). Preliminary results for HCHO and BrO from the EOS-aura ozone monitoring instrument. In S. C. Tsay, T. Yokota, & M.-H. Ahn (Eds.), *Passive optical remote sensing of the atmosphere and clouds IV* (Vol. 5652, pp. 116–123). SPIE. https://doi.org/10.1117/12.578606

Kuttippurath, J., Abbhishek, K., Gopikrishnan, G., & Pathak, M. (2022). Investigation of long–term trends and major sources of atmospheric HCHO over India. *Environmental Challenges*, *7*, 100477. https://doi.org/10.1016/j.envc.2022.100477

Kwon, H.-A., Abad, G. G., Nowlan, C., Chong, H., Souri, A., Vigouroux, C., et al. (2023). Validation of OMPS Suomi NPP and OMPS NOAA-20 formaldehyde total columns with NDACC FTIR observations. *Earth and Space Science*, *10*(5), e2022EA002778. https://doi.org/10.1029/2022ea002778

Kwon, H.-A., Park, R. J., González Abad, G., Chance, K., Kurosu, T. P., Kim, J., et al. (2019). Description of a formaldehyde retrieval algorithm for the geostationary environment monitoring spectrometer (GEMS). *Atmospheric Measurement Techniques*, *12*(7), 3551–3571. https://doi.org/10.5194/amt-12-3551-2019

Kwon, H.-A., Park, R. J., Jeong, J. I., Lee, S., González Abad, G., Kurosu, T. P., et al. (2017). Sensitivity of formaldehyde (HCHO) column measurements from a geostationary satellite to temporal variation of the air mass factor in East Asia. *Atmospheric Chemistry and Physics*, *17*(7), 4673–4686. https://doi.org/10.5194/acp-17-4673-2017

Langerock, B., De Mazière, M., Hendrick, F., Vigouroux, C., Desmet, F., Dils, B., & Niemeijer, S. (2015). Description of algorithms for co-locating and comparing gridded model data with remote-sensing observations. *Geoscientific Model Development*, *8*(3), 911–921. https://doi.org/10.5194/gmd-8-911-2015

Lee, G. T., Park, R. J., Kwon, H.-A., Ha, E. S., Lee, S. D., Shin, S., et al. (2024). First evaluation of the GEMS formaldehyde product against TROPOMI and ground-based column measurements during the in-orbit test period. *Atmospheric Chemistry and Physics*, *24*(8), 4733–4749. https://doi.org/10.5194/acp-24-4733-2024

Levelt, P. F., Joiner, J., Tamminen, J., Veefkind, J. P., Bhartia, P. K., Stein Zweers, D. C., et al. (2018). The ozone monitoring instrument: Overview of 14 years in space. *Atmospheric Chemistry and Physics*, *18*(8), 5699–5745. https://doi.org/10.5194/acp-18-5699-2018

Levelt, P. F., Van Den Oord, G. H., Dobber, M. R., Malkki, A., Visser, H., De Vries, J., et al. (2006). The ozone monitoring instrument. *IEEE Transactions on Geoscience and Remote Sensing*, *44*(5), 1093–1101. https://doi.org/10.1109/tgrs.2006.872333

Levitus, S., & US NODC, U. N. O. D. C. (2013). NODC standard product: World Ocean Atlas 2009 (NCEI Accession 0094866) [Dataset]. *NOAA National Centers for Environmental Information*. https://www.ncei.noaa.gov/access/metadata/landing-page/bin/iso?id=gov.noaa.nodc:0094866

Li, C., Joiner, J., Krotkov, N. A., & Dunlap, L. (2015). A new method for global retrievals of HCHO total columns from the Suomi national polar-orbiting partnership ozone mapping and profiler suite. *Geophysical Research Letters*, *42*(7), 2515–2522. https://doi.org/10.1002/2015gl063204

Li, D., Wang, S., Xue, R., Zhu, J., Zhang, S., Sun, Z., & Zhou, B. (2021). OMI-observed HCHO in shanghai, China, during 2010–2019 and ozone sensitivity inferred by an improved HCHO/NO_2 ratio. *Atmospheric Chemistry and Physics*, *21*(20), 15447–15460. https://doi.org/10.5194/acp-21-15447-2021

Liao, J., Hanisco, T. F., Wolfe, G. M., St Clair, J., Jimenez, J. L., Campuzano-Jost, P., et al. (2019). Towards a satellite formaldehyde–in situ hybrid estimate for organic aerosol abundance. *Atmospheric Chemistry and Physics*, *19*(5), 2765–2785. https://doi.org/10.5194/acp-19-2765-2019

Lloyd, S. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, *28*(2), 129–137. https://doi.org/10.1109/TIT.1982.1056489

Malicet, J., Daumont, D., Charbonnier, J., Parisse, C., Chakir, A., & Brion, J. (1995). Ozone UV spectroscopy. II. Absorption cross-sections and temperature dependence. *Journal of Atmospheric Chemistry*, *21*(3), 263–273. https://doi.org/10.1007/bf00696758

Marais, E. A., Jacob, D. J., Guenther, A., Chance, K., Kurosu, T. P., Murphy, J. G., et al. (2014). Improved model of isoprene emissions in Africa using ozone monitoring instrument (OMI) satellite observations of formaldehyde: Implications for oxidants and particulate matter. *Atmospheric Chemistry and Physics*, *14*(15), 7693–7703. https://doi.org/10.5194/acp-14-7693-2014

Marais, E. A., Jacob, D. J., Jimenez, J. L., Campuzano-Jost, P., Day, D. A., Hu, W., et al. (2016). Aqueous-phase mechanism for secondary organic aerosol formation from isoprene: Application to the Southeast United States and co-benefit of SO_2 emission controls. *Atmospheric Chemistry and Physics*, *16*(3), 1603–1618. https://doi.org/10.5194/acp-16-1603-2016

Marais, E. A., Jacob, D. J., Kurosu, T., Chance, K., Murphy, J., Reeves, C., et al. (2012). Isoprene emissions in Africa inferred from OMI observations of formaldehyde columns. *Atmospheric Chemistry and Physics*, *12*(14), 6219–6235. https://doi.org/10.5194/acp-12-6219-2012

Martin, R. V., Chance, K., Jacob, D. J., Kurosu, T. P., Spurr, R. J., Bucsela, E., et al. (2002). An improved retrieval of tropospheric nitrogen dioxide from GOME. *Journal of Geophysical Research*, *107*(D20), ACH9-1–ACH9-21. https://doi.org/10.1029/2001jd001027

Martin, R. V., Fiore, A. M., & Van Donkelaar, A. (2004). Space-based diagnosis of surface ozone sensitivity to anthropogenic emissions. *Geophysical Research Letters*, *31*(6), L06120. https://doi.org/10.1029/2004gl019416

Millet, D. B., Jacob, D. J., Boersma, K. F., Fu, T.-M., Kurosu, T. P., Chance, K., et al. (2008). Spatial distribution of isoprene emissions from North America derived from formaldehyde column measurements by the OMI satellite sensor. *Journal of Geophysical Research*, *113*(D2), D02307. https://doi.org/10.1029/2007jd008950

NASA Aura. (2024). Systems analysis: Potential evolution of the aura mission. Retrieved from https://aura.gsfc.nasa.gov/Potential_Evolution_of_Aura.html

NASA NSPIRES. (2022). NASA's terra, aqua, and aura drifting orbits workshop November 1-2, 2022. Retrieved from https://nspires.nasaprs.com/external/viewrepositorydocument/cmdocumentid=929460/solicitationId=%7B19F4296E-5280-3996-3149-42CB166328DC%7D/viewSolicitationDocument=1/TAA_ExecutiveSummary.pdf

NDACC. (2023). Network for the detection of atmospheric composition change (NDACC) public data access [Dataset]. *National Aeronautics and Space Administration (NASA)*. Retrieved from https://www-air.larc.nasa.gov/missions/ndacc/data.html

NOAA. (1976). *US standard atmosphere*. National Oceanic and Atmospheric Administration.

Nowlan, C. R., González Abad, G., Kwon, H.-A., Ayazpour, Z., Chan Miller, C., Chance, K., et al. (2023). Global formaldehyde products from the ozone mapping and profiler suite (OMPS) nadir mappers on Suomi NPP and NOAA-20. *Earth and Space Science*, *10*(5), e2022EA002643. https://doi.org/10.1029/2022ea002643

Palmer, P. I., Jacob, D. J., Chance, K., Martin, R. V., Spurr, R. J., Kurosu, T. P., et al. (2001). Air mass factor formulation for spectroscopic measurements from satellites: Application to formaldehyde retrievals from the global ozone monitoring experiment. *Journal of Geophysical Research*, *106*(D13), 14539–14550. https://doi.org/10.1029/2000jd900772

Richter, A., Begoin, M., Hilboll, A., & Burrows, J. (2011). An improved NO_2 retrieval for the GOME-2 satellite instrument. *Atmospheric Measurement Techniques*, *4*(6), 1147–1159. https://doi.org/10.5194/amt-4-1147-2011

Rodgers, C. D., & Connor, B. J. (2003). Intercomparison of remote sounding instruments. *Journal of Geophysical Research*, *108*(D3), 4116. https://doi.org/10.1029/2002jd002299

Schaaf, C., & Wang, Z. (2015). MCD43C1 MODIS/Terra+Aqua BRDF/AlbedoModel parameters daily L3 global 0.05Deg CMG V006 [Dataset]. *Distributed by NASA EOSDIS Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MODIS/MCD43C1.006

Schaaf, C., & Wang, Z. (2021). MODIS/Terra+Aqua BRDF/AlbedoModel parameters daily L3 global 0.05Deg CMG V061 [Dataset]. *Distributed by NASA EOSDIS Land Processes Distributed Active Archive Center*. https://doi.org/10.5067/MODIS/MCD43C1.061

Schenkeveld, V., Jaross, G., Marchenko, S., Haffner, D., Kleipool, Q. L., Rozemeijer, N. C., et al. (2017). In-flight performance of the ozone monitoring instrument. *Atmospheric Measurement Techniques*, *10*(5), 1957–1986. https://doi.org/10.5194/amt-10-1957-2017

Serdyuchenko, A., Gorshelev, V., Weber, M., Chehade, W., & Burrows, J. P. (2014). High spectral resolution ozone absorption cross-sections–part 2: Temperature dependence. *Atmospheric Measurement Techniques*, *7*(2), 625–636. https://doi.org/10.5194/amt-7-625-2014

Souri, A. H., Choi, Y., Jeon, W., Woo, J.-H., Zhang, Q., & Kurokawa, J.-I. (2017). Remote sensing evidence of decadal changes in major tropospheric ozone precursors over East Asia. *Journal of Geophysical Research: Atmospheres*, *122*(4), 2474–2492. https://doi.org/10.1002/2016JD025663

Souri, A. H., Nowlan, C. R., González Abad, G., Zhu, L., Blake, D. R., Fried, A., et al. (2020). An inversion of NO_x and non-methane volatile organic compound (NMVOC) emissions using satellite observations during the KORUS-AQ campaign and implications for surface ozone over East Asia. *Atmospheric Chemistry and Physics*, *20*(16), 9837–9854. https://doi.org/10.5194/acp-20-9837-2020

Spurr, R. J. (2006). VLIDORT: A linearized pseudo-spherical vector discrete ordinate radiative transfer code for forward model and retrieval studies in multilayer multiple scattering media. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *102*(2), 316–342. https://doi.org/10.1016/j.jqsrt.2006.05.005

Spurr, R. J. (2008). LIDORT and VLIDORT: Linearized pseudo-spherical scalar and vector discrete ordinate radiative transfer models for use in remote sensing retrieval problems. In *Light scattering reviews 3: Light scattering and reflection* (pp. 229–275).

Spurr, R. J., & Christi, M. (2019). The LIDORT and VLIDORT linearized scalar and vector discrete ordinate radiative transfer models: Updates in the last 10 years. In *Springer series in light scattering: Volume 3: Radiative transfer and light scattering* (pp. 1–62).

Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., Van Der Werf, G., Giglio, L., & Guenther, A. (2009a). Evaluating the performance of pyrogenic and biogenic emission inventories against one decade of space-based formaldehyde columns. *Atmospheric Chemistry and Physics*, *9*(3), 1037–1060. https://doi.org/10.5194/acp-9-1037-2009

Stavrakou, T., Müller, J.-F., De Smedt, I., Van Roozendael, M., Van Der Werf, G., Giglio, L., & Guenther, A. (2009b). Global emissions of non-methane hydrocarbons deduced from SCIAMACHY formaldehyde columns through 2003–2006. *Atmospheric Chemistry and Physics*, *9*(11), 3663–3679. https://doi.org/10.5194/acp-9-3663-2009

Su, W., Liu, C., Hu, Q., Zhang, C., Liu, H., Xia, C., et al. (2022). First global observation of tropospheric formaldehyde from Chinese GaoFen-5 satellite: Locating source of volatile organic compounds. *Environmental Pollution*, *297*, 118691. https://doi.org/10.1016/j.envpol.2021.118691

Su, W., Liu, C., Hu, Q., Zhao, S., Sun, Y., Wang, W., et al. (2019). Primary and secondary sources of ambient formaldehyde in the Yangtze River delta based on ozone mapping and profiler suite (OMPS) observations. *Atmospheric Chemistry and Physics*, *19*(10), 6717–6736. https://doi.org/10.5194/acp-19-6717-2019

Sun, K., Liu, X., Huang, G., González Abad, G., Cai, Z., Chance, K., & Yang, K. (2017). Deriving the slit functions from OMI solar observations and its implications for ozone-profile retrieval. *Atmospheric Measurement Techniques*, *10*(10), 3677–3695. https://doi.org/10.5194/amt-10-3677-2017

Sun, K., Zhu, L., Cady-Pereira, K., Chan Miller, C., Chance, K., Clarisse, L., et al. (2018). A physics-based approach to oversample multi-satellite, multispecies observations to a common grid. *Atmospheric Measurement Techniques*, *11*(12), 6679–6701. https://doi.org/10.5194/amt-11-6679-2018

Surl, L., Palmer, P. I., & González Abad, G. (2018). Which processes drive observed variations of HCHO columns over India? *Atmospheric Chemistry and Physics*, *18*(7), 4549–4566. https://doi.org/10.5194/acp-18-4549-2018

Thalman, R., & Volkamer, R. (2013). Temperature dependent absorption cross-sections of O_2–O_2 collision pairs between 340 and 630 nm and at atmospherically relevant pressure. *Physical Chemistry Chemical Physics*, *15*(37), 15371–15381. https://doi.org/10.1039/c3cp50968k

Thomas, W., Hegels, E., Slijkhuis, S., Spurr, R. J., & Chance, K. (1998). Detection of biomass burning combustion products in Southeast Asia from backscatter data taken by the GOME spectrometer. *Geophysical Research Letters*, *25*(9), 1317–1320. https://doi.org/10.1029/98gl01087

Tilstra, L. G., Tuinder, O. N. E., Wang, P., & Stammes, P. (2017). Surface reflectivity climatologies from UV to NIR determined from Earth observations by GOME-2 and SCIAMACHY. *Journal of Geophysical Research: Atmospheres*, *122*(7), 4084–4111. https://doi.org/10.1002/2016JD025940

Travis, K., Judd, L., Crawford, J., Chen, G., Szykman, J., Whitehill, A., et al. (2022). Can column formaldehyde observations inform air quality monitoring strategies for ozone and related photochemical oxidants? *Journal of Geophysical Research: Atmospheres*, *127*(13), e2022JD036638. https://doi.org/10.1029/2022jd036638

U.S. National Ice Center. (2008). IMS daily Northern Hemisphere snow and ice analysis at 1 km, 4 km, and 24 km resolutions, version 1 [Dataset]. *National Snow Ice Data Center*. https://doi.org/10.7265/N52R3PMC

Valin, L., Fiore, A., Chance, K., & González Abad, G. (2016). The role of OH production in interpreting the variability of CH_2O columns in the Southeast US. *Journal of Geophysical Research: Atmospheres*, *121*(1), 478–493. https://doi.org/10.1002/2015jd024012

Vandaele, A. C., Hermans, C., Simon, P. C., Carleer, M., Colin, R., Fally, S., et al. (1998). Measurements of the NO_2 absorption cross-section from 42 000 cm- 1 to 10 000 cm- 1 (238–1000 nm) at 220 K and 294 K. *Journal of Quantitative Spectroscopy and Radiative Transfer*, *59*(3–5), 171–184. https://doi.org/10.1016/s0022-4073(97)00168-4

Veefkind, J. P., Boersma, K. F., Wang, J., Kurosu, T. P., Krotkov, N., Chance, K., & Levelt, P. F. (2011). Global satellite analysis of the relation between aerosols and short-lived trace gases. *Atmospheric Chemistry and Physics*, *11*(3), 1255–1267. https://doi.org/10.5194/acp-11-1255-2011

Veefkind, J. P., de Haan, J. F., Sneep, M., & Levelt, P. F. (2016). Improvements to the OMI O_2–O_2 operational cloud algorithm and comparisons with ground-based radar–lidar observations. *Atmospheric Measurement Techniques*, *9*(12), 6035–6049. https://doi.org/10.5194/amt-9-6035-2016

Vigouroux, C., Bauer Aquino, C. A., Bauwens, M., Becker, C., Blumenstock, T., De Mazière, M., et al. (2018). NDACC harmonized formaldehyde time series from 21 FTIR stations covering a wide range of column abundances. *Atmospheric Measurement Techniques*, *11*(9), 5049–5073. https://doi.org/10.5194/amt-11-5049-2018

Vigouroux, C., Langerock, B., Bauer Aquino, C. A., Blumenstock, T., Cheng, Z., De Mazière, M., et al. (2020). TROPOMI–Sentinel-5 precursor formaldehyde validation using an extensive network of ground-based Fourier-transform infrared stations. *Atmospheric Measurement Techniques*, *13*(7), 3751–3767. https://doi.org/10.5194/amt-13-3751-2020

Vrekoussis, M., Wittrock, F., Richter, A., & Burrows, J. P. (2010). GOME-2 observations of oxygenated VOCs: What can we learn from the ratio glyoxal to formaldehyde on a global scale? *Atmospheric Chemistry and Physics*, *10*(21), 10145–10160. https://doi.org/10.5194/acp-10-10145-2010

Wang, H., González Abad, G., Chan Miller, C., Kwon, H.-A., Nowlan, C. R., Ayazpour, Z., et al. (2023). Development of the measures blue band water vapor algorithm—Towards a long-term data record. *Atmospheric Measurement Techniques Discussions*, 1–32. https://doi.org/10.5194/amt-2023-66

Wang, H., Wu, Q., Guenther, A. B., Yang, X., Wang, L., Xiao, T., et al. (2021). A long-term estimation of biogenic volatile organic compound (BVOC) emission in China from 2001–2016: The roles of land cover change and climate variability. *Atmospheric Chemistry and Physics*, *21*(6), 4825–4848. https://doi.org/10.5194/acp-21-4825-2021

Wang, Z., Schaaf, C. B., Sun, Q., Shuai, Y., & Román, M. O. (2018). Capturing rapid land surface dynamics with Collection V006 MODIS BRDF/NBAR/Albedo (MCD43) products. *Remote Sensing of Environment*, *207*, 50–64. https://doi.org/10.1016/j.rse.2018.02.001

Werdell, J., O'Reilly, J., Hu, C., Feng, L., Lee, Z., Franz, B., et al. (2023). Chlorophyll a, NASA algorithm publication tool, 2023-11-06, v1.1. https://doi.org/10.5067/JCQB8QALDOYD

Wilmouth, D. M., Hanisco, T. F., Donahue, N. M., & Anderson, J. G. (1999). Fourier transform ultraviolet spectroscopy of the A $^2\pi_{3/2} \leftarrow X^2\pi_{3/2}$ transition of BrO. *The Journal of Physical Chemistry A*, *103*(45), 8935–8945. https://doi.org/10.1021/jp991651o

Wittrock, F., Richter, A., Oetjen, H., Burrows, J. P., Kanakidou, M., Myriokefalitakis, S., et al. (2006). Simultaneous global observations of glyoxal and formaldehyde from space. *Geophysical Research Letters*, *33*(16), L16804. https://doi.org/10.1029/2006GL026310

Wolfe, G. M., Nicely, J. M., Clair, J. M. S., Hanisco, T. F., Liao, J., Oman, L. D., et al. (2019). Mapping hydroxyl variability throughout the global remote troposphere via synthesis of airborne and satellite formaldehyde observations. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(23), 11171–11180. https://doi.org/10.1073/pnas.1821661116

Wu, X., Wen, J., Xiao, Q., You, D., Liu, Q., & Lin, X. (2018). Forward a spatio-temporal trend surface for long-term ground-measured albedo upscaling over heterogeneous land surface. *International Journal of Digital Earth*, *11*(5), 470–484. https://doi.org/10.1080/17538947.2017.1334097

Zhao, T., Mao, J., Ayazpour, Z., González Abad, G., Nowlan, C. R., & Zheng, Y. (2024). Interannual variability of summertime formaldehyde (HCHO) vertical column density and its main drivers at northern high latitudes. *Atmospheric Chemistry and Physics*, *24*(10), 6105–6121. https://doi.org/10.5194/acp-24-6105-2024

Zhou, Y., Brunner, D., Boersma, K. F., Dirksen, R., & Wang, P. (2009). An improved tropospheric NO_2 retrieval for OMI observations in the vicinity of mountainous terrain. *Atmospheric Measurement Techniques*, *2*(2), 401–416. https://doi.org/10.5194/amt-2-401-2009

Zhu, L., González Abad, G., Nowlan, C. R., Chan Miller, C., Chance, K., Apel, E. C., et al. (2020). Validation of satellite formaldehyde (HCHO) retrievals using observations from 12 aircraft campaigns. *Atmospheric Chemistry and Physics*, *20*(20), 12329–12345. https://doi.org/10.5194/acp-20-12329-2020

Zhu, L., Jacob, D. J., Keutsch, F. N., Mickley, L. J., Scheffe, R., Strum, M., et al. (2017). Formaldehyde (HCHO) as a hazardous air pollutant: Mapping surface air concentrations from satellite and inferring cancer risks in the United States. *Environmental Science & Technology*, *51*(10), 5650–5657. https://doi.org/10.1021/acs.est.7b01356

Zhu, L., Jacob, D. J., Kim, P. S., Fisher, J. A., Yu, K., Travis, K. R., et al. (2016). Observing atmospheric formaldehyde (HCHO) from space: Validation and intercomparison of six retrievals from satellites (OMI, GOME2a, GOME2b, OMPS) with seac^{4}rs aircraft observations over the Southeast US. *Atmospheric Chemistry and Physics*, *16*(21), 13477–13490. https://doi.org/10.5194/acp-16-13477-2016

Zhu, L., Mickley, L. J., Jacob, D. J., Marais, E. A., Sheng, J., Hu, L., et al. (2017). Long-term (2005–2014) trends in formaldehyde (HCHO) columns across North America as seen by the OMI satellite instrument: Evidence of changing emissions of volatile organic compounds. *Geophysical Research Letters*, *44*(13), 7079–7086. https://doi.org/10.1002/2017GL073859

Zoogman, P., Liu, X., Suleiman, R., Pennington, W., Flittner, D., Al-Saadi, J., et al. (2017). Tropospheric emissions: Monitoring of pollution (TEMPO). *Journal of Quantitative Spectroscopy and Radiative Transfer*, *186*, 17–39. https://doi.org/10.1016/j.jqsrt.2016.05.008